Automatic Monte-Carlo Tuning for Minimum Bias Events at the LHC

DISSERTATION

zur Erlangung des akademischen Grades

Doctor Rerum Naturalium
(Dr. rer. nat.)
im Fach Physik

eingereicht an der
Mathematisch-Naturwissenschaftlichen
Fakultät I
Humboldt-Universität zu Berlin

von
Sami Kama
geboren am 18.09.1978 in Biga/Türkiye

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Dr. h.c. Christoph Markschies

Dekan der Mathematisch-Naturwissenschaftlichen
Fakultät I:
Prof. Dr. Lutz-Helmut Schön

Gutachter:
1. Prof. Dr. Hermann Kolanoski
2. Dr. Klaus Mönig
3. Prof. Dr. Torbjörn Sjöstrand

eingereicht am: 24.03.2010
Tag der mündlichen Prüfung: 22.06.2010

To my Angel.

Abstract

The Large Hadron Collider near Geneva Switzerland will ultimately collide protons at a center-of-mass energy of 14 TeV and 40 MHz bunch crossing rate with a luminosity of $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. At each bunch crossing about 20 soft proton-proton interactions are expected to happen. In order to study new phenomena and improve our current knowledge of the physics these events must be understood. However, the physics of soft interactions are not completely known at such high energies. Different phenomenological models, trying to explain these interactions, are implemented in several Monte-Carlo (MC) programs such as PYTHIA, PHOJET and EPOS. Some parameters in such MC programs can be tuned to improve the agreement with the data.

In this thesis a new method for tuning the MC programs, based on Genetic Algorithms and distributed analysis techniques have been presented. This method represents the first and fully automated MC tuning technique that is based on true MC distributions. It is an alternative to parametrization-based automatic tuning.

This new method is used in finding new tunes for PYTHIA 6 and 8. These tunes are compared to the tunes found by alternative methods, such as the PROFESSOR framework and manual tuning, and found to be equivalent or better. Charged particle multiplicity, $dN_{\mathrm{ch}}/d\eta$, Lorentz-invariant yield, transverse momentum and mean transverse momentum distributions at various center-of-mass energies are generated using default tunes of EPOS, PHOJET and the Genetic Algorithm tunes of PYTHIA 6 and 8. These distributions are compared to measurements from UA5, CDF, CMS and ATLAS in order to investigate the best model available. Their predictions for the ATLAS detector at LHC energies have been investigated both with generator level and full detector simulation studies.

Comparison with the data did not favor any model implemented in the generators, but EPOS is found to describe investigated distributions better. New data from ATLAS and CMS show higher than expected multiplicities and a faster rise with the center-of-mass energy in central particle multiplicity.

Zusammenfassung

Der “Large Hadron Collider” am CERN bei Genf in der Schweiz ist gebaut worden, um Protonen bei einer Schwerpunktsenergie von 14 TeV mit einer Strahlkreuzungsrate von 40 MHz bei einer Luminosität von $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ kollidieren zu lassen. Bei jeder Strahlkreuzung enstehen dann ca. 20 weiche Proton-Proton-Wechselwirkungen, deren überlagerte Signale vom Detektor gemessen werden. Diese Ereignisse müssen so präzise wie möglich verstanden werden, um einerseits neuartige physikalische Phänomene zu entdecken, andererseits aber dazu beitragen, das Verständnis bereits bestehender physikalischer Gesetze zu verbessern. Inbesondere ist die Physik der weichen Wechselwirkungen momentan noch nicht genau verstanden. Unterschiedliche theoretische Modelle, die versuchen, diese Physik zu beschreiben, sind in Monte-Carlo (MC) Generatoren wie EPOS, PHOJET und PYTHIA eingebunden. Deren Modelle sind auf mannigfaltige Weise parametrisierbar und müssen mit experimentellen Daten angepasst werden.

In dieser Arbeit wird eine neue Methode, basierend auf genetischen Algorithmen und verteilten Analysetechniken, vorgestellt, um diese MC-Parameter anzupassen. Diese Methode stellt einen alternativen Ansatz zu derzeit verfügbaren Methode wie PROFESSOR dar mit dem Vorteil, dass die Suche nach geeigneten Modellparametern automatisiert ist.

Der Ansatz der genetischen Algorithmen wurde benutzt, um für PYTHIA 6 und PYTHIA 8 Parameter zu finden, die gut mit bisherigen Messungen übereinstimmen. Die Ergebnisse wurden mit den MC-Generatoren EPOS und PHOJET und Daten von UA5 CDF, CMS und ATLAS verglichen, wobei eine Reihe von charakteristischen Verteilungen untersucht wurden wie Multiplizitäts- Spektren geladener Teilchen sowie Lozentz-invariante Größen. Auch Vorhersagen für LHC-Energien werden sowohl auf Generator level als auch nach kompletter ATLAS-Detektor-Simulation präsentiert.

Datenvergleiche beveoruzugen nicht eindeutig eines der in die Generatoren implementierten Modelle, jedoch beschreibt EPOS die untersuchten Verteilungen etwas besser. Neue Daten von ATLAS und CMS zeigen höhere Multiplizitäten als erwartet und einen schnelleren Anstieg der zentralen Multiplizität mit der Schwerpunktsenergie.

Contents

1. Introduction

Large Hadron Collider (LHC) that has been built near Geneva will be a large step in testing of the Standard Model of particle physics and extend the search of new physics even further. There are four major experiments at the LHC: ALICE, ATLAS, CMS, and LHCb. ATLAS and CMS are general purpose detectors, on the other hand LHCb is oriented towards b-quark studies and ALICE is designed for heavy-ion collisions. The LHC, at its nominal design energy of 14 TeV and luminosity $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, is expected to have 720-920 million interactions per second. Most of these interactions will be soft QCD interactions also known as Minimum Bias (MB) events. The physics of the minimum bias events is not completely understood, yet they are the key for understanding the detectors and search for new physics. It is estimated that at each bunch crossing at nominal energy and luminosity about 20 minimum bias interactions will take place [1, 2].

Most of the events have to be filtered out since it is impossible to store that much data and only a very small fraction of these events have interesting signatures for new physics. Thus one needs to understand the minimum bias events for searching interesting event signatures and selecting the data to be stored for offline processing. Moreover MB events will affect any measurement significantly. Figure 1.1 shows a simulation of a typical $Z \to j + j$ event at the ATLAS detector with and without MB pileup.

The traditional solution for these problems is using Monte-Carlo (MC) programs for estimating the trigger settings and event selection criteria (cuts) for the background subtraction. However since the soft QCD interactions are not yet completely understood, there are different MC programs with different models. The models in such programs usually agree up to some extent with the minimum bias data however their extrapolations to the higher energies are usually quite different. Also these models usually have free parameters that needs to be tuned in order to be able to describe existing data better. Tuning of the MC programs has been a tedious and subjective task. In this thesis a new method using genetic algorithms and distributed processing is presented for automatic MC tuning.

In chapter 2, a brief introduction to the LHC and the ATLAS experiment is presented. The data acquisition and trigger systems are discussed and a brief explanation of trigger monitoring is given in chapter 3. Chapter 4 contains a brief introduction to the Standard Model, Monte-Carlo programs and their models. In chapter 5, the Genetic Algorithms and the tuning method is discussed. The data sets used in tuning are also discussed there. Chapter 6 contains the comparisons between data and MC. Chapter 7 presents the predictions of different generators at LHC energies. The conclusions are given in the chapter 8. Three appendices contain the detailed description of the Operational Monitoring Display, various data-MC and MC-MC comparison plots.

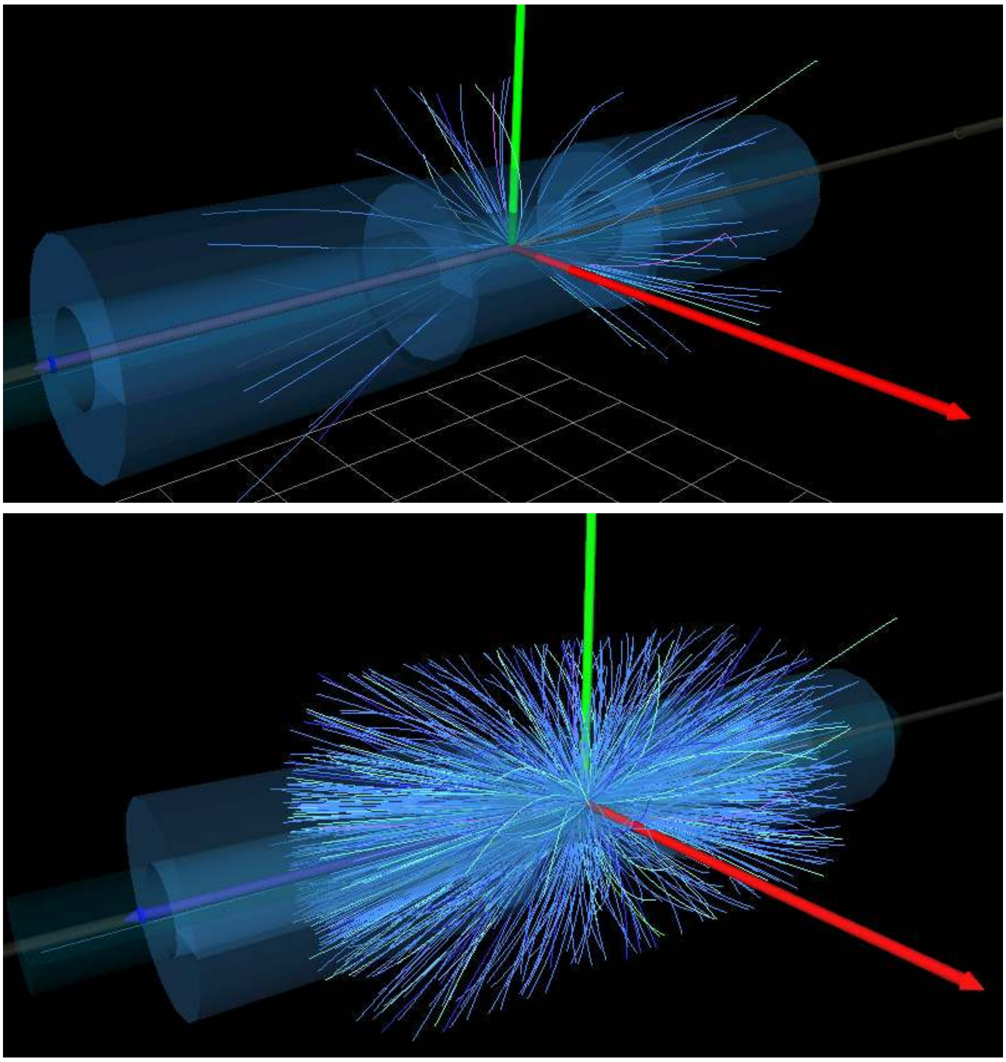


Figure 1.1.: Simulation of a typical $Z^0 \rightarrow j + j$ event in the ATLAS detector. Top picture shows only the tracks from the event and the bottom picture shows the tracks from the same event together with the tracks from 23 minimum bias events that are expected to happen in the same bunch crossing at the LHC.

2. LHC and ATLAS

The Standard Model and beyond will be subjected to precision tests at the Large Hadron Collider (LHC) up to a center-of-mass energy of 14 TeV [3]. The LHC has been built underground near the Swiss-French border, at the European Nuclear Research Center (CERN) near Geneva. Construction of the LHC machine was approved by the CERN council in December 1996. In 2000, the Large Electron Positron (LEP) collider was closed and construction of the LHC began. The LHC started operation in September 2008 but a faulty connection caused an accident and it was shutdown until Nowember 2009 for repairs.

At the LHC there are two high luminosity experiments, ATLAS [4] and CMS [5], and two low luminosity experiments LHCb [6] and TOTEM [7]. There is also one dedicated heavy-ion experiment, ALICE [8]. Their locations on the LHC ring and accelerator complex at CERN can be seen in figure 2.1. LHCb is oriented towards b quark physics, with a peak luminosity of $\mathcal{L} = 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. TOTEM, on the other hand, looks for elastic scattering at small angles, with a peak luminosity of $\mathcal{L} = 2 \cdot 10^{29}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. ALICE aims to study heavy-ion (lead-lead) collisions at a peak luminosity of $\mathcal{L} = 10^{27}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. ATLAS and CMS are both general purpose experiments designed to search for physics of the Standard Model and beyond. Only the ATLAS experiment is described here. Detailed descriptions of the other experiments and their physics goals are available in the respective references.

2.1. The Large Hadron Collider

The Large Hadron Collider (LHC) is a two-ring superconducting hadron accelerator and collider built in the existing 26.7 km tunnel constructed for LEP. The plane of the tunnel lies between 45 m and 170 m below the surface and is inclined at a 1.4% slope towards Lac Léman. Since the tunnel from LEP has been reused, the LHC machine is constrained by the size of the tunnel. Because of this constraint, the machine uses a “two-in-one” superconducting magnet design. Proton beams in the machine rotate in separate pipes with separate magnetic fields and vacuum chambers except in the insertion and experimental detector regions.

The nominal design center-of-mass energy and peak luminosity of the LHC are 14 TeV and $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, respectively. To reach 14 TeV, the dipole magnets must generate a nominal magnetic field of 8.33 T. Design luminosity is achieved by injecting 2808 bunches, seperated by 25 ns, with $1.15 \cdot 10^{11}$ maximum protons per bunch. These bunches are accelerated in several different stages of accelerators. The accelerator stages and routes of different beams are shown in figure 2.1. Bunches with 450 GeV are injected into the LHC in both directions. It can take from 70 minutes to 7 hours for the machine

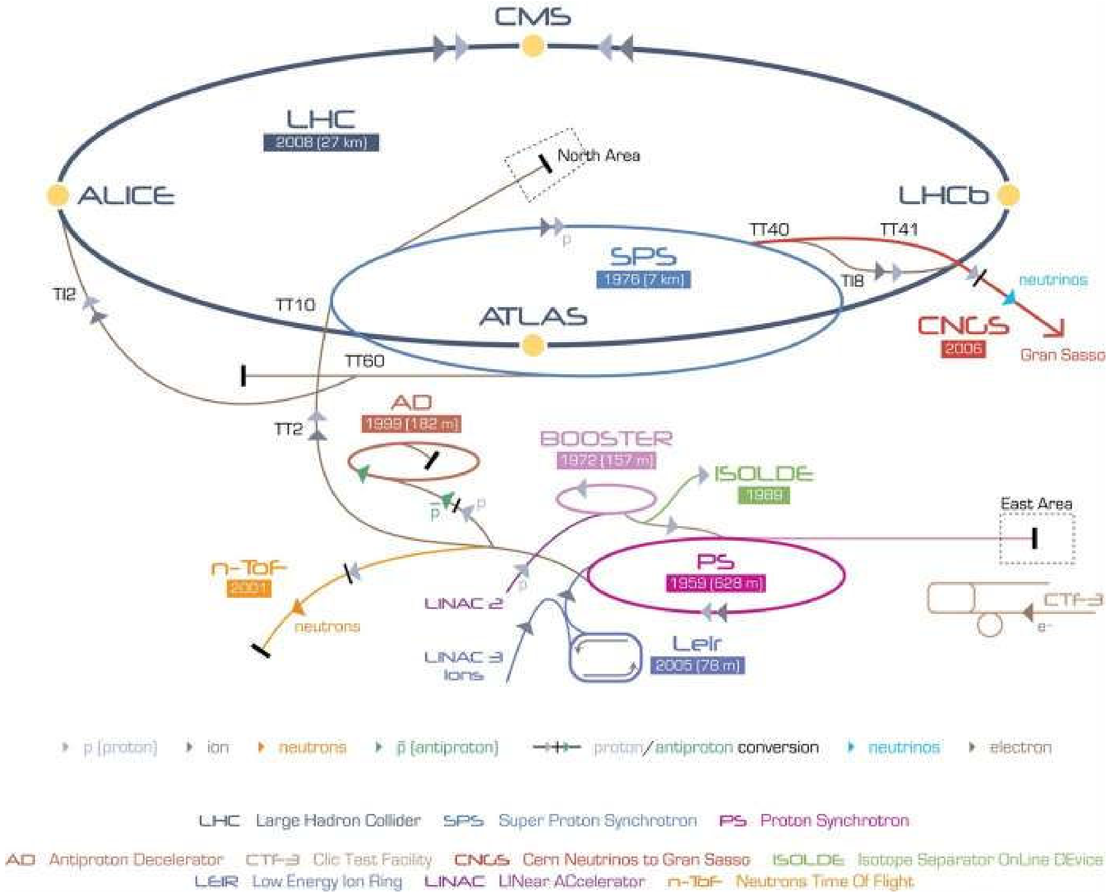


Figure 2.1.: Accelerator complex at CERN. Both proton bunches are accelerated to 450 GeV in several accelerator stages before being injected into the LHC, where they are accelerated up to 7 TeV. The locations of the four major experiments at the LHC ring are also shown.

to be ready for collisions [3]. Once the machine is ready, it is expected to run for 15 hours before the intensity drops below an acceptable limit. It is predicted that the LHC will have a maximum total integrated luminosity between $80\,\mathrm{fb}^{-1}$ and $120\,\mathrm{fb}^{-1}$ per year.

At peak operation, with a total beam current of 0.584 A, the LHC will have about 1 GJ of energy stored in the beams and magnets. In case of emergency, the beams will be dumped into a graphite target and the magnet systems will be forcibly quenched by quench heaters. Magnetic energy is dispersed with heaters and dump resistors.

2.2. ATLAS Detector

The ATLAS experiment is one of the two general purpose experiments at the LHC. ATLAS is an acronym for **A T**oroidal **L**HC **A**pparatu**S**. The ATLAS Collaboration aims to measure the Standard Model parameters and search for new physics phenomena. Precise tracking and calorimetry information are needed to accomplish these physics goals requiring a state-of-the-art detector. The ATLAS detector has been built to meet these requirements with the performance goals listed in table 2.1. It has a cylindrical shape and is composed of several layers of sub-detectors. As shown in figure 2.2, the layout of the detector from inside to outside is: Pixel Detector, Silicon Microstrip Detector, Transition Radiation Tracker, Solenoid Magnet, Liquid Argon Electromagnetic Calorimeters and Tile Calorimeters, Toroidal Magnets, Monitored Drift Tubes, Resistive Plate Chambers and Thin Gap Chambers. The first three form the Inner Detector, the second three form the calorimetry systems and the last three form the muon systems.

Detector Component	**Required resolution**	**η coverage**	
		Measurement	Trigger
Tracking	$\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 0.05\% \cdot p_\mathrm{T} \oplus 1\%$	± 2.5	
EM Calorimetry	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	± 3.2	± 2.5
Hadronic Calorimetry(jets)			
barrel and endcap	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	± 3.2	± 3.2
forward	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$	$3.1 < \lvert\eta\rvert < 4.9$	$3.1 < \lvert\eta\rvert < 4.9$
Muon spectrometer	$\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 10\%$ at $p_\mathrm{T} = 1$ TeV	± 2.7	± 2.4

Table 2.1.: ATLAS detector performance goals. The symbol $\oplus$ represents addition in quadrature.

ATLAS uses the nominal interaction point to define a right-handed coordinate system. The z-direction of the coordinate system lies on the LHC beam line such that the positive x-direction points to the center of the LHC ring and the positive y-direction points upwards. Transverse quantities including transverse momentum p_T, transverse energy E_T and missing transverse energy $E_\mathrm{T}^\mathrm{miss}$, are defined in the x-y plane i.e. $p_\mathrm{T} = \sqrt{p_x^2 + p_y^2}$. The azimuthal angle ϕ is defined around the z-axis and the polar angle θ is defined as the

the opening angle from the $+z$-axis. The pseudo-rapidity, η, is defined as $\eta = -\ln\tan\theta/2$ and distance in the η–ϕ plane is defined as $\Delta\mathrm{R} = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

In the following sections of this chapter, the parts of the ATLAS detector are briefly described. Details of the detector can be found in [1, 2, 4] and the references therein.

2.2.1. Inner Detector

The Inner Detector produces high precision measurements of charged particle tracks. In order to achieve the physics goals of the experiment it is designed to provide excellent momentum resolution for both primary and secondary vertex measurements of charged particles down to $p_\mathrm{T} = 100$ MeV within the pseudo-rapidity range $|\eta| < 2.5$. Figure 2.3 shows one-quarter of the sub-detectors and their distances from the nominal interaction point. The Inner Detector resides in a 2 T solenoidal magnetic field and has a cylindrical shape with boundaries ±3512 mm in length and 1150 mm in radius. It is the innermost part of the detector, surrounding the beam pipe and therefore is the most irradiated component of the detector. The design of the Inner Detector is constrained by the requirement of high precision, limits of existing technology, extreme running conditions and by the costs. In order to meet the requirements, it has been designed in three independent but complementary parts. The innermost part, the Pixel Detector, provides high resolution space points. The middle part, the SemiConductor Tracker (SCT) is a silicon strip detector which complements the pixel space points with stereo pairs of silicon microstrip layers. The outermost Inner Detector layer, the Transition Radiation Tracker (TRT), provides continuous tracking to enhance pattern recognition, thereby improving the momentum resolution, as well as electron and pion identification complementary to that of the calorimeters. Figure 2.4 shows the cut-out view of the Inner Detector.

Due to the extreme conditions near the interaction point, the innermost layer of the Pixel Detector will need to be replaced after three years of operation at design luminosity. In order to keep the noise levels manageable, the silicon sensors are kept at temperatures between $-5°$ C to $-10°$ C which corresponds to coolant temperatures of $\sim -25°$ C. However the TRT operates at room temperatures. Thus the mechanical structure has been designed to able to withstand such temperature gradients with minimal distortions.

Pixel Detector

The Pixel Detector is composed of 1744 pixel sensors. The sensor modules are placed on three cylindrical layers, layer 0, 1 and 2, in the barrel region and on three disks in the endcap regions on both sides. The barrel layers extend from $z = -400.5$ mm to $z = 400.5$ mm and reside at radial distances of $R = 50.5$, 88.5 and 122.5 mm. The disks are located at $z = \pm495$, ±580 and ±650 mm covering the radial range $88.5 < R < 149.6$ mm. This setup of the layers and disks provides spacepoints for charged particles within $|\eta| < 2.5$.

Each sensor module has a size of 19 x 63 mm^2. They will initially be operated at $\sim$ 150 V bias voltage which can be increased up to 600 V throughout the lifetime of the experiment. 90% of the chips have pixels with dimensions of 50 x 400 μm^2 and the

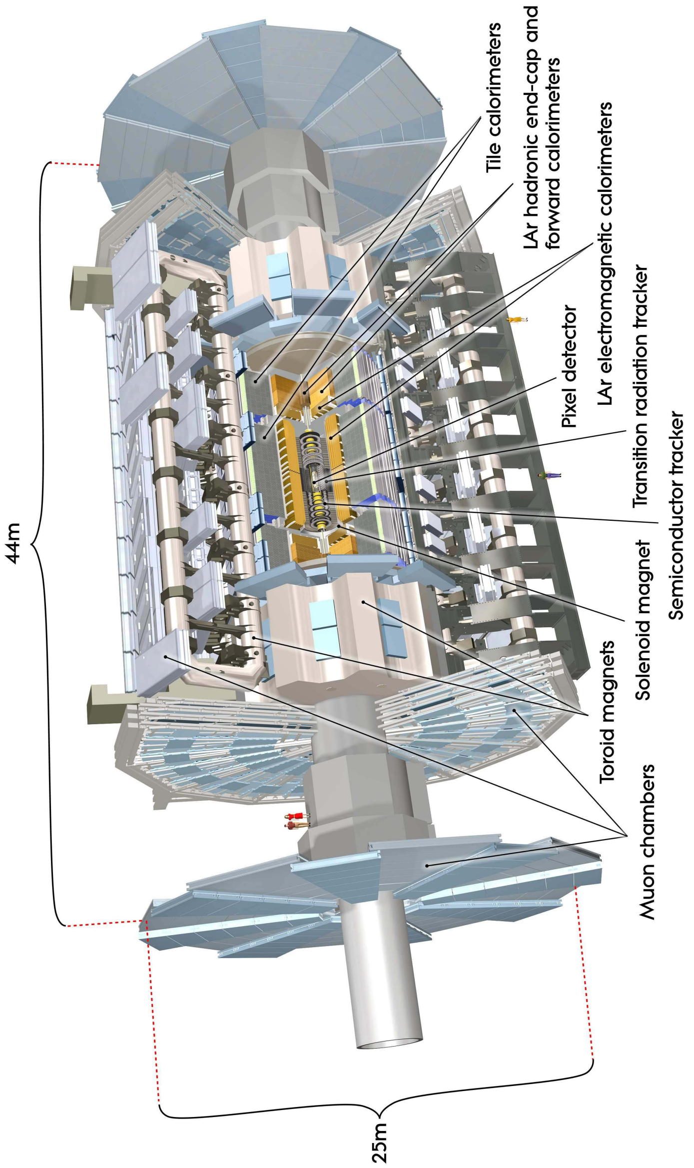


Figure 2.2.: The ATLAS detector with its sub-detectors. It has a diameter of 25 m, a width of 44 m and weighs 7000 tonnes.

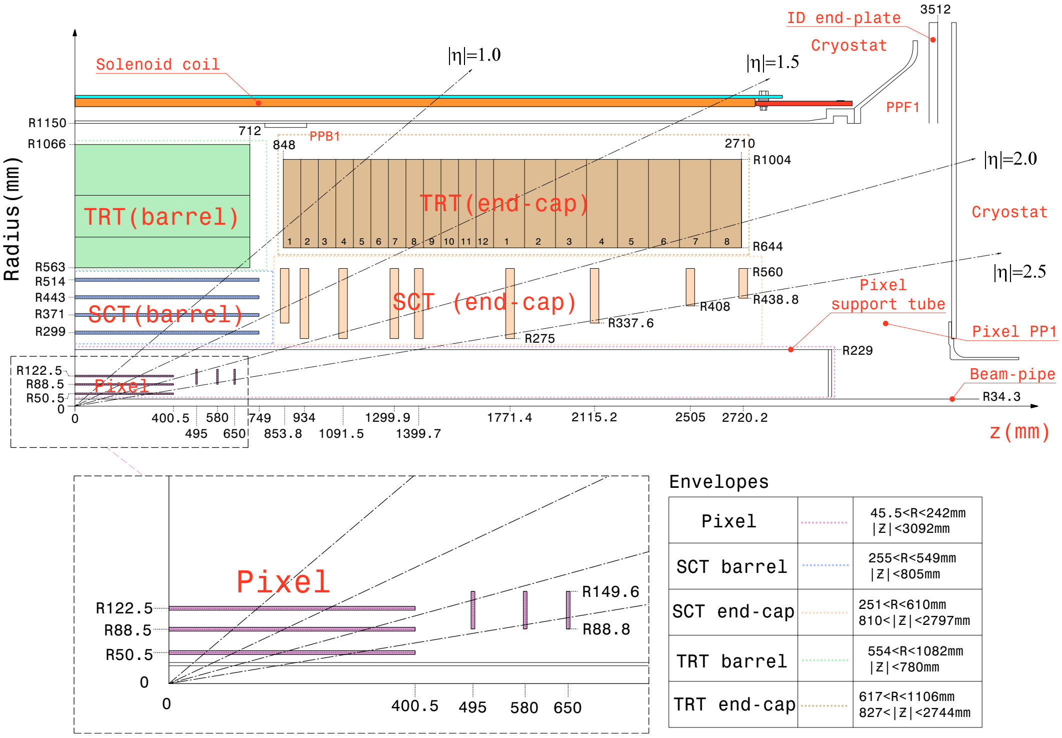


Figure 2.3.: One quarter of the sub-detectors of the Inner Detector with their distances from the nominal interaction point in mm. The top picture shows the barrel and endcap regions of the TRT, SCT and Pixel Detector. The bottom picture shows a close-up of the Pixel Detector layout.

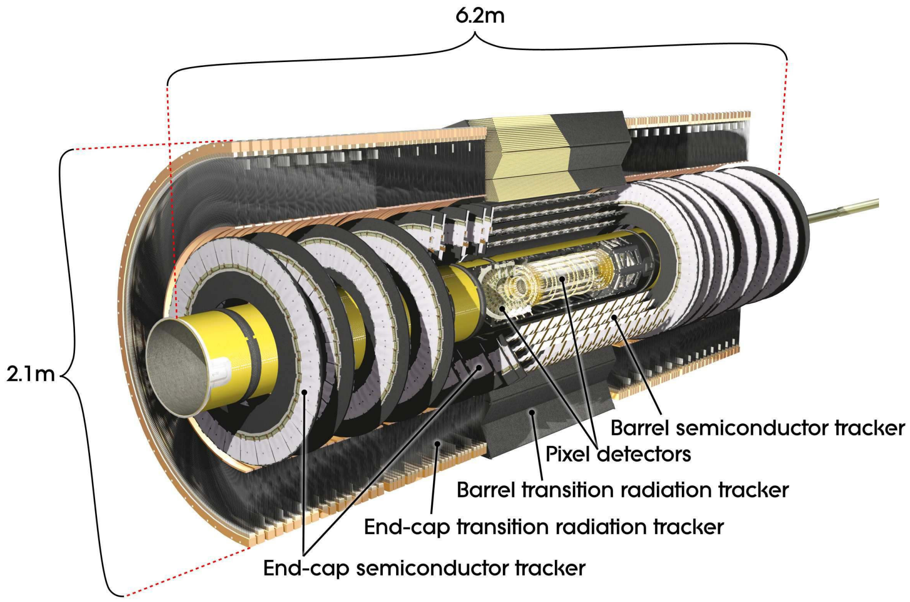


Figure 2.4.: A 3D rendered image of the Inner Detector. The Pixel Detector, in the center, is surrounded by the SCT barrel layers. TRT straws are placed after SCT layers. Pixel and SCT endcap disks are visible on both sides. Note that SCT endcap disks are surrounded by TRT endcap straws.

remaining part have $50 \times 600\,\mu\mathrm{m}^2$. Although each sensor module contains 47232 pixels, due to space constraints there are only 46080 readout channels per module. Therefore the number of total readout channels in the Pixel Detector is approximately 80.4 million.

To measure the momentum of the particles accurately, the locations of the pixels must be precisely known. In order to attain an intrinsic accuracy of $10\,\mu$m in the R–ϕ and $115\,\mu$m in z-directions, the position of each module must be known within $20\,\mu$m in z and $7\,\mu$m in R–ϕ directions. The radial uncertainty of the modules should be less than $10\,\mu$m for layer-0 and $20\,\mu$m for layer-1 and layer-2.

Silicon Microstrip Detector

The Silicon Microstrip Detector (SCT) is the second sub-detector of the inner detector. It is situated after the pixel layers. It contains 4088 modules; 2112 of which are in four coaxial cylindrical layers in the barrel region and the remaining 1976 are in eighteen disk layers; nine in each endcap region. The modules are made of single-sided p-in-n type silicon strips, paired with a pitch of $80\,\mu$m and glued back-to-back with a 40 mrad angle at the center points; providing stereoscopic measurements. The initial bias voltage of the modules is 150 V and will be increased up to 350 V during the lifetime of the experiment.

Transition Radiation Tracker

The Transition Radiation Tracker (TRT) is composed of 4 mm diameter straw drift tubes made of two $35\,\mu$m thick multi-layer films bonded back-to-back. The straws located in the barrel region are 144 cm long, and in the endcap regions they are 37 cm in length.

The position of the anode wire is one of the essential parameters for an accurate measurement. The anode wire is supported mechanically by a plastic insert glued to the inner wall of the straw in order to keep the anode wire offset from the straw center within $300\,\mu$m. This creates an inefficient region at the center of the straw. Such straws have an effective, active length of 71.2 cm on both sides. The straws are typically operated at approximately 1530 V and filled with a 70% Xe, 27% CO_2 and 3% O_2 mixture at 5–10 mbar over-pressure. The Xe-based gas mixture is continuously monitored and circulated inside the straws to maintain operation quality. Moreover the straws are kept in a CO_2 envelope at room temperature to prevent contamination. The gas mixture in the straws has been specifically selected to maximise the efficiency of the photon absorption from the transition radiation photons emitted by the electrons passing through the straws. The signals generated by the transition radiation are digitized by comparing them to a low and a high threshold; encoding the information about the pulse shape. This information is used in electron identification, making the TRT an electron discriminator as well as a tracker.

Tracking performance

ATLAS track parametrization uses the azimuthal angle ϕ, polar angle $\cot\theta$, transverse impact parameter d_0, the longitudinal impact parameter $z_0 \cdot \sin\theta$ and the charge over transverse momentum q/p_{T} to represent each track at its perigee. θ is defined with respect to the z-axis and ϕ is defined on the x-y plane, with respect to the x-axis. z_0 is the longitudinal distance from the interaction point and d_0 is defined as the transverse distance from the z-axis. The sign of d_0 is positive when the track direction is clockwise with respect to the origin, negative otherwise.

The resolutions of the track parameters depend on the track p_{T} and η and can be expressed as [9]

$$\sigma_{\mathrm{X}}(p_{\mathrm{T}}) = \sigma_{\mathrm{X}}(\infty)(1 \oplus \frac{p_{\mathrm{x}}}{p_{\mathrm{T}}}) \tag{2.1}$$

where X represents a track parameter, $\sigma_{\mathrm{X}}(\infty)$ is the asymptotic resolution expected at infinite momentum and p_{x} is the constant momentum at which the intrinsic detector resolution is equal to the resolution uncertainity contribution due to multiple-scattering. The $\oplus$ sign represents addition in quadrature.

Figure 2.5 shows the resolutions of the tracking parameters with respect to track η for minimum bias events obtained from full detector simulation. Figure 2.6 shows the resolutions with respect to track p_{T}. Figure 2.7 shows the impact parameter resolutions of the muons and pions with 5 GeV in the range $|\eta| \leq 0.5$ [9]. Details of the tracking performance can be found in [9, 10].

2.2.2. Calorimetry

The ATLAS detector contains electromagnetic (EM) and hadronic (HAD) calorimetry systems covering the range $|\eta| < 4.9$. Figure 2.8 shows the types and the coverages of the calorimeters in the detector. The Liquid Argon (LAr) EM calorimeter is divided into a barrel and two endcap components. The barrel region matching the Inner Detector has a finer granularity than other parts of the calorimeter to enable precision measurements of electrons and photons. At $\eta = 0$, the thickness of the EM calorimeter is greater than 22 radiation lengths (X_0) and the total active calorimeter is approximately 9.7 interaction lengths (λ). Together with the outer support structure, the total interaction length becomes $11\,\lambda$, effectively reducing the punch-through to well below the irreducible level of prompt or decay muons. The large η-coverage and long interaction lengths of the calorimeters ensure a good E_{T} measurement which is necessary for many important physics processes.

Electromagnetic Calorimeter

The EM calorimeter is a lead-LAr detector with accordion-shaped kapton electrodes and lead absorber plates, providing complete ϕ symmetry without any azimuthal cracks. The barrel part, covering $|\eta| < 1.475$, and the two endcap parts, within $1.375 < |\eta| < 3.2$, are each housed in their own cryostat. To minimize the material in front of the calorimeter, the central LAr and the central Solenoid Magnet around the Inner Detector share a common vacuum vessel.

The endcap calorimeters are divided into two coaxial wheels, the outer providing coverage in $1.375 < |\eta| < 2.5$ and the inner in $2.5 < |\eta| < 3.2$. The barrel calorimeter is composed of two half barrels with a 4 mm gap in between them at z = 0. There is a

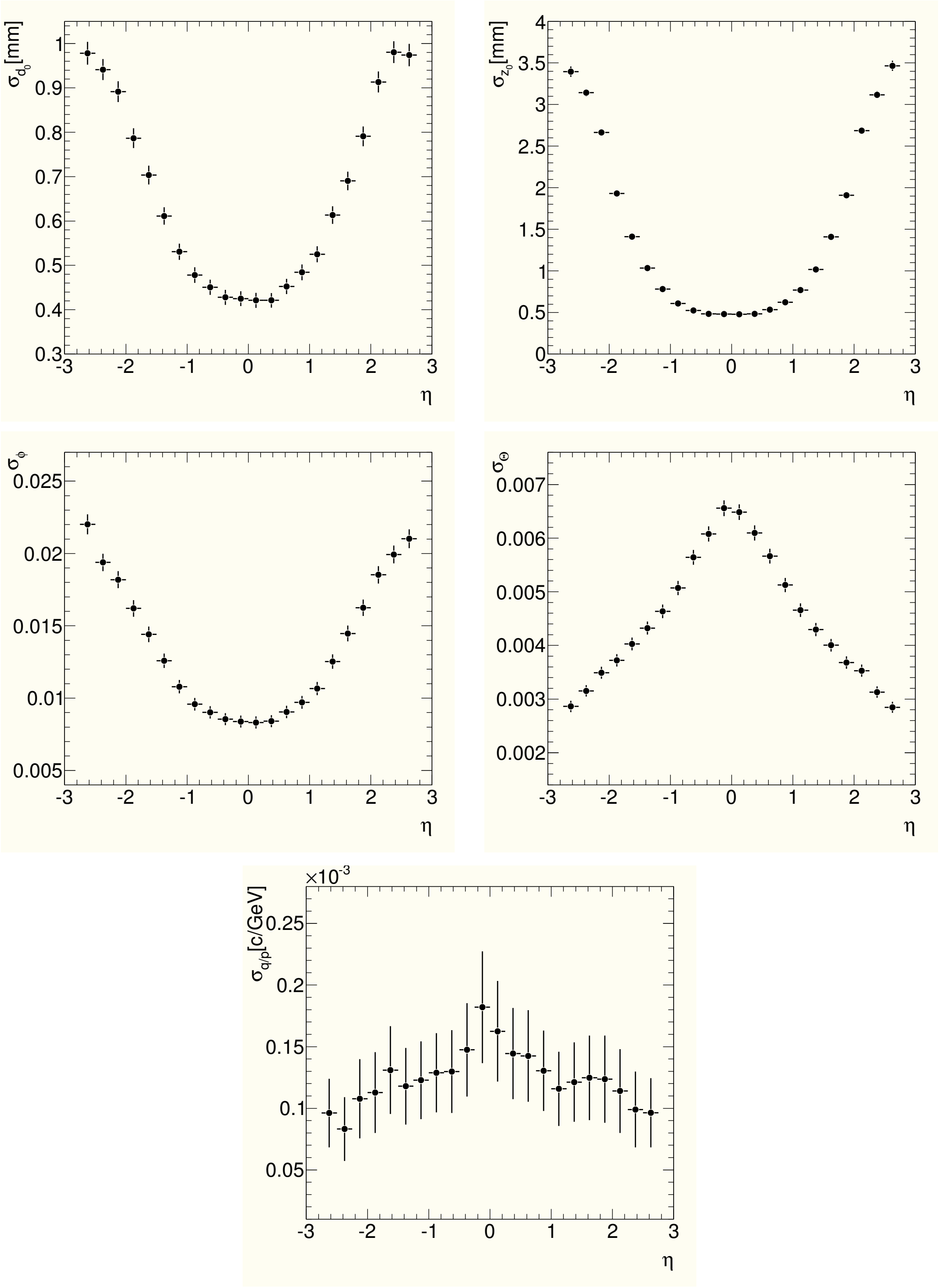


Figure 2.5.: Pseudorapidity dependence of the track parameter resolutions of charged particles with $p_{\mathrm{T}} > 0.15$ GeV in minimum bias events. The top two plots show d_0 and z_0 on the left and on the right, respectively. The middle row shows ϕ and θ. The bottom row shows q/p. Taken from [10].

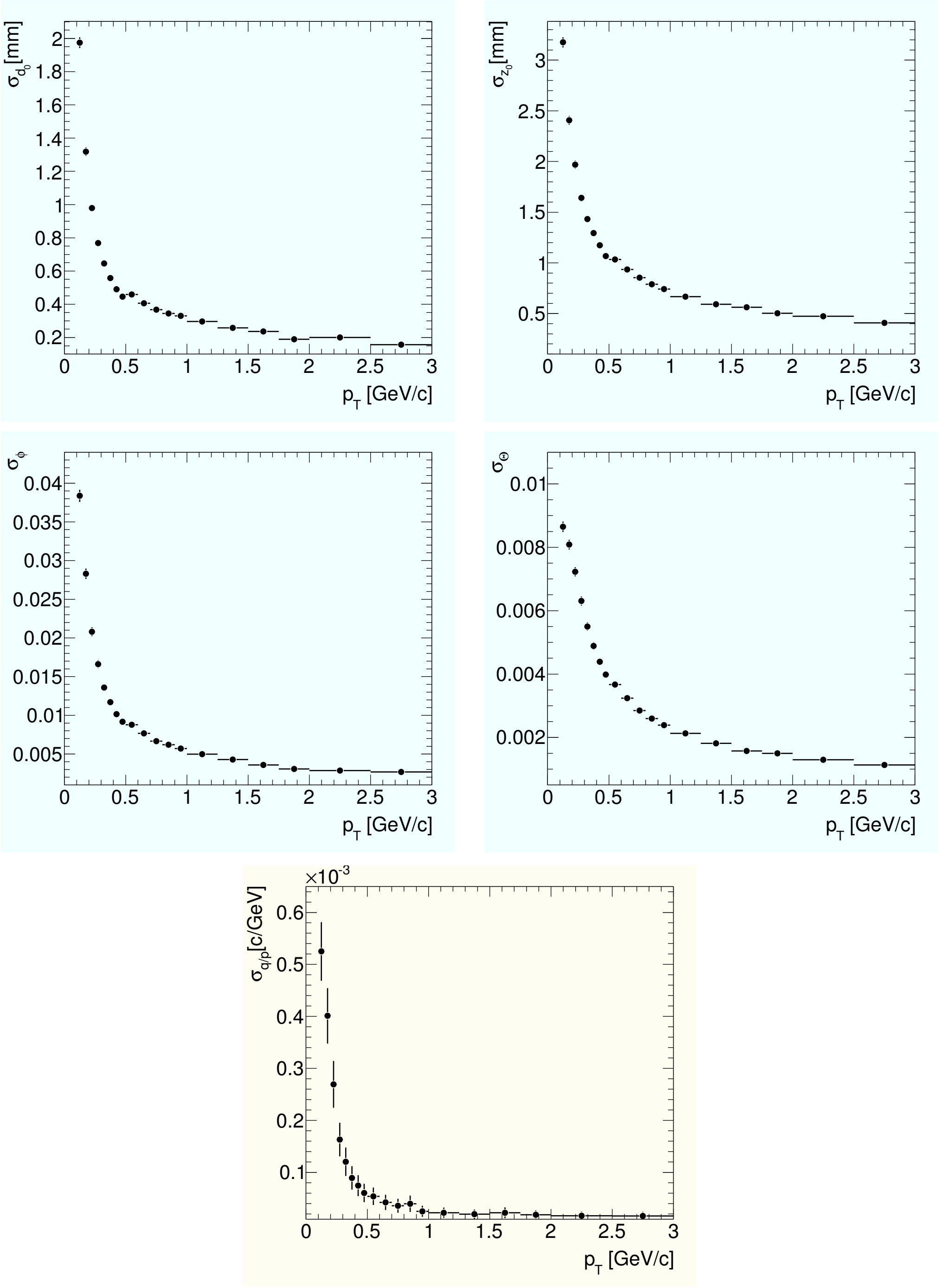


Figure 2.6.: Transverse momentum dependence of the track parameter resolutions of charged particles with $p_{\mathrm{T}} > 0.15$ GeV in minimum bias events. Top two plots show d_0 and z_0 on the left and on the right, respectively. Middle row shows ϕ and θ. Bottom row shows q/p. Taken from [10].

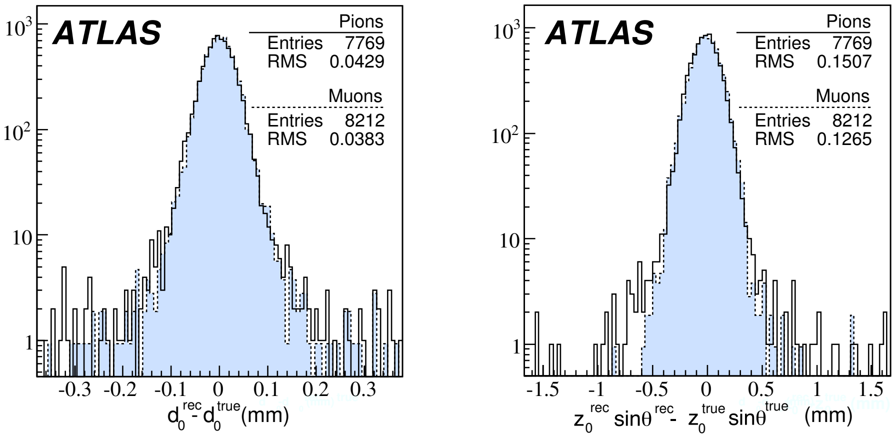


Figure 2.7.: Impact parameter resolution distributions of the muons and pions with $p_T = 5$ GeV in $|\eta| \leq 0.5$. d_0 is on the left and the $z_0 \times \sin\theta$ is on the right.

presampler layer in front of the calorimeter to increase the electron detection efficiency.

The presampler layer has a granularity of $\Delta\eta \times \Delta\phi = 0.025 \times 0.1$. In the barrel part, the granularity of the calorimeter changes from 0.025×0.1 to 0.050×0.025, depending on the radius and position. The granularity of the endcap calorimeters changes from 0.050×0.1 to 0.1×0.1. Figure 2.9 shows a module of the LAr EM barrel.

Hadronic Calorimeters

There are three different types of hadronic calorimeter used in the ATLAS detector. The Tile calorimeter resides directly above the LAr EM calorimeter and covers $|\eta| < 1.7$ together with its barrel and two extended barrel parts. It is made with steel absorbers and scintillating tiles as the active material. The LAr Hadronic Endcap Calorimeters (HEC) consist of two independent wheels per endcap and they are placed inside the LAr EM cryostats. The HEC overlaps with the Tile calorimeter and the Forward Calorimeter (FCal) and is made up of wedge-shaped modules. The modules are made of Copper plates as the absorber and liquid Argon as the active medium. The Forward Calorimeter is integrated into the endcap cryostat and has approximately $10\,\lambda$ interaction lengths. It is made with Copper and Tungsten as the absorbers and liquid Argon as the active medium.

The granularity of the Tile Calorimeter is $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$, except for the last layer where it is 0.2×0.1 for both the barrel and the endcap regions. The LAr HEC has a granularity of 0.1×0.1 in $1.5 < |\eta| < 2.5$ and 0.2×0.2 in $2.5 < |\eta| < 3.2$. The FCal has granularity values ranging from $\Delta x \times \Delta y = 3.0 \times 2.6\,\text{cm}$ to $5.4 \times 4.7\,\text{cm}$. The details of the calorimeters are available in [4, 11] and references therein.

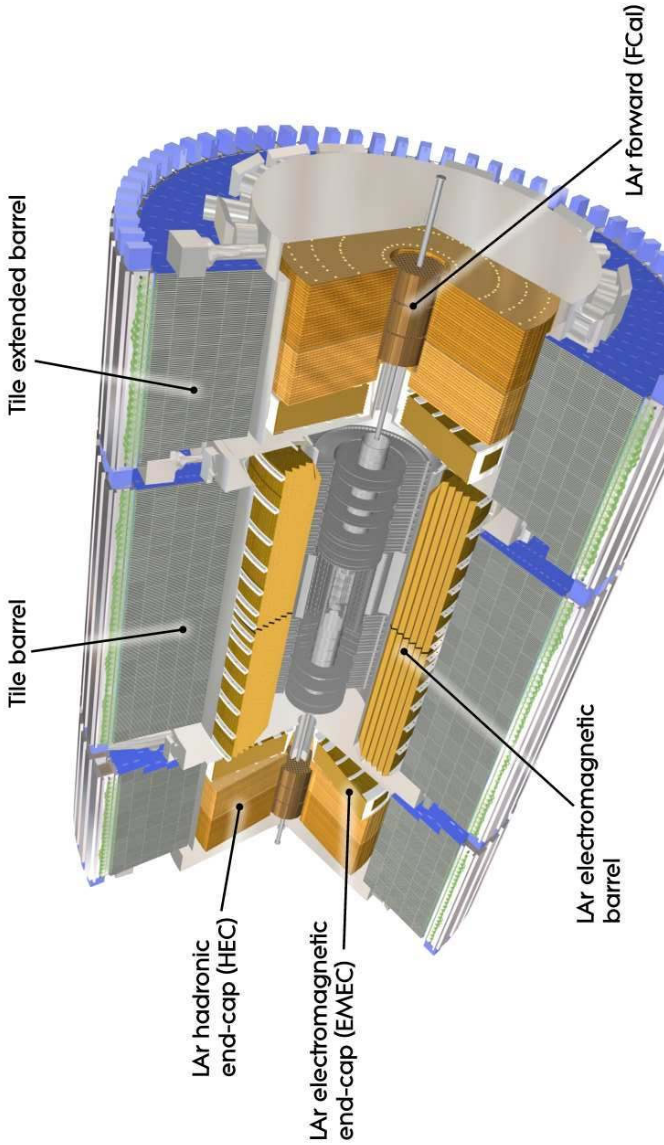


Figure 2.8.: Atlas calorimeter systems surrounding the Inner Detector. The outermost layer is the Tile Calorimeter. The Liquid Argon (LAr) Electromagnetic (EM) Calorimeter resides inside the Tile Calorimeter. Both endcap regions contain the LAr EM and Hadronic as well as the Forward Calorimeters.

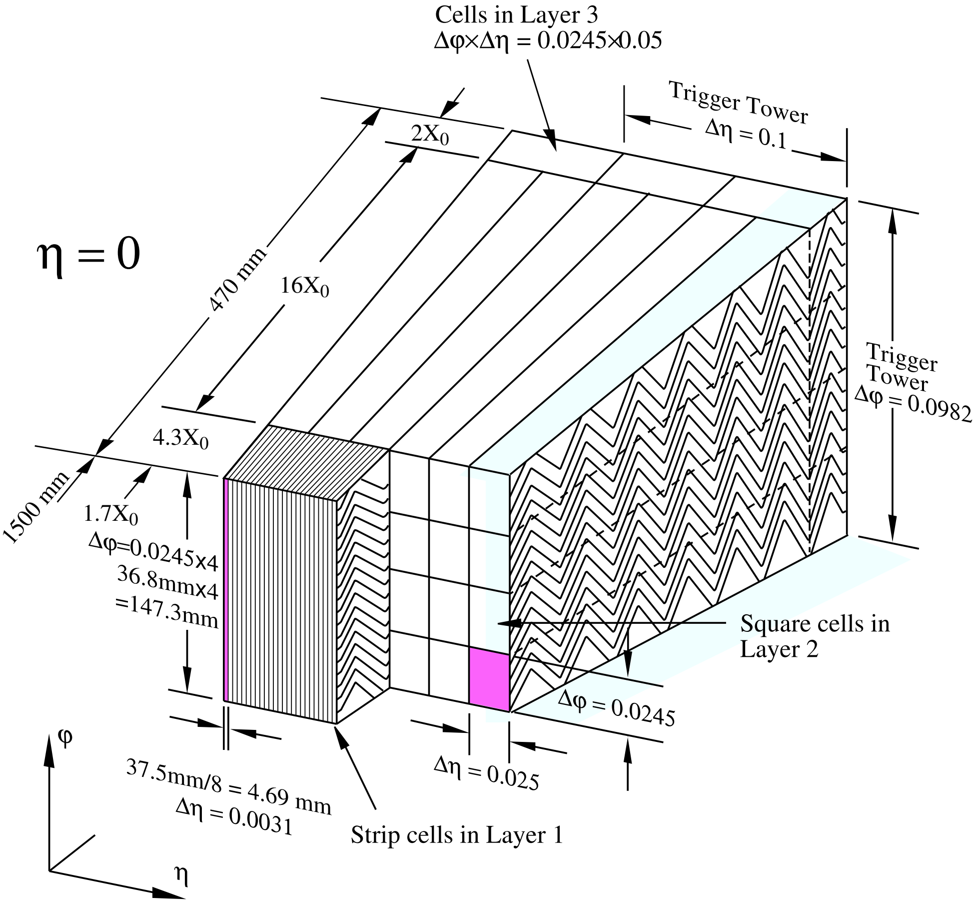


Figure 2.9.: Part of a barrel LAr EM module.

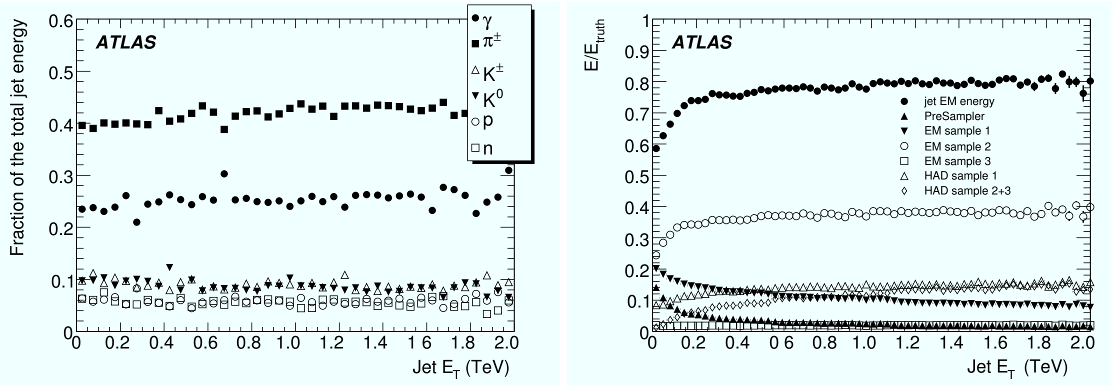


Figure 2.10.: Jet energy fraction carried by different particles and the energy deposited in different calorimeter layers. Energy fractions are calculated by applying the jet reconstruction algorithms to the generator level particles. Taken from [9].

Jets in Calorimeters

Jet reconstruction efficiencies and resolutions in ATLAS are extensively studied using full detector simulations [9]. Figure 2.10 shows the particle compositon of the jet energy and the energy deposited in the different parts of the calorimeters. Although the particle composition of the jets is mostly independent of the jet energy; the reconstructed jets are affected by the detector properties and bending by magnetic fields. Reconstructed jets are corrected for various effects to obtain the best estimator for the true jet energy. “True jet energy” is defined by applying jet reconstruction algorithms to the generator level particles. Figure 2.11 shows the ratio of the reconstructed energy to the true jet energy at two different η values and jet energies. Figure 2.12 shows the true jet energy dependence of the ratio $E_{\text{rec}}/E_{\text{true}}$ for two different algorithms at two different η ranges. Figure 2.13 shows the η dependency of the ratio of the reconstructed and true transverse energy of the jets, $E_T^{\text{rec}}/E_T^{\text{true}}$, for two different jet reconstruction algorithms at three different energy ranges. Jet reconstruction algorithms, efficiencies, corrections and other details are available in [9] and the references therein.

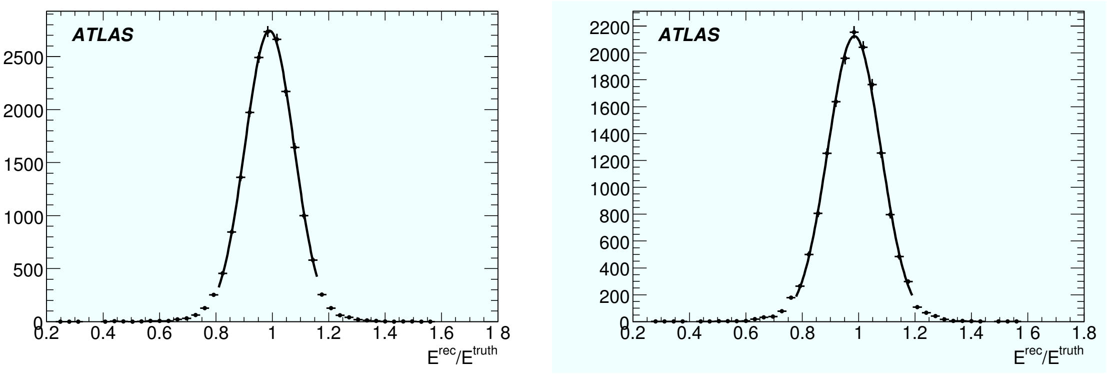


Figure 2.11.: Ratio of reconstructed jet energy to true jet energy. The left plot shows the ratio for the jets in the range $88 < E_{\text{truth}} < 107$ GeV and $|\eta| < 0.5$ and the right plot show the ratio for $158 < E_{\text{truth}} < 191$ GeV and $1.0 < |\eta| < 1.5$. Solid line is a gaussian fit. Taken from [9].

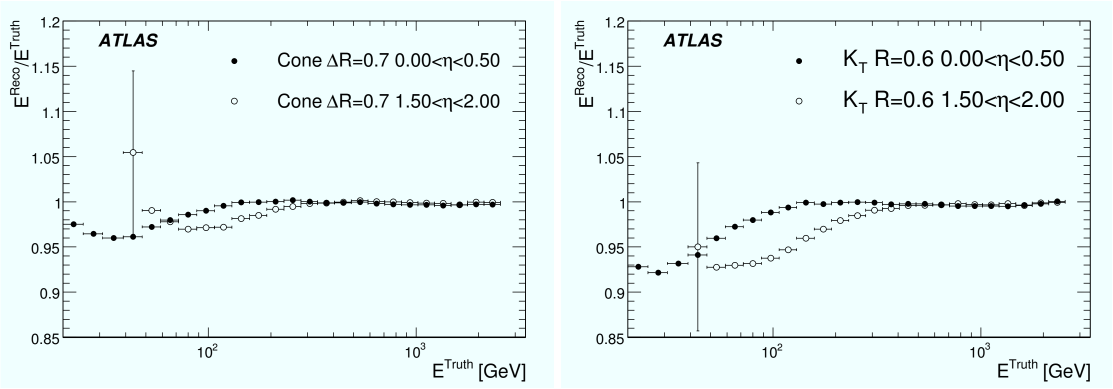


Figure 2.12.: Ratio of the reconstructed jet energy to the true jet energy with respect to true jet energy for two different jet reconstruction algorithms at two different η. The left plot shows the ratio for the Cone algorithm with $R_{\text{cone}} = 0.7$ in the range $|\eta| < 0.5$ (black) and $1.5 < |\eta| < 2.0$ (white). The right plot shows the ratio for the same ranges using the k_T algorithm with $R = 0.6$. Taken from [9].

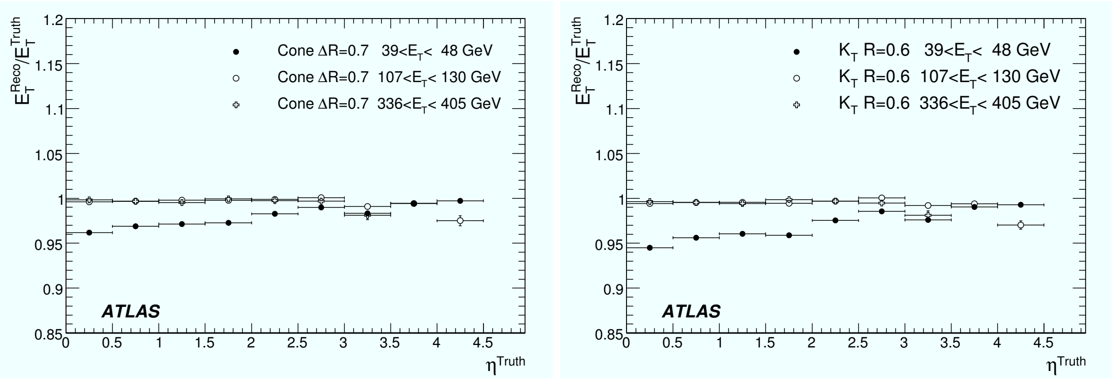


Figure 2.13.: The η dependence of $E_T^{\text{rec}}/E_T^{\text{true}}$ of the Cone algorithm and the k_t algorithm for three different energy ranges. Left plot shows the η dependence of the Cone algorithm with $R_{\text{cone}} = 0.7$ and the right plot shows the ratio for the k_T algorithm with $R = 0.6$. Taken from [9].

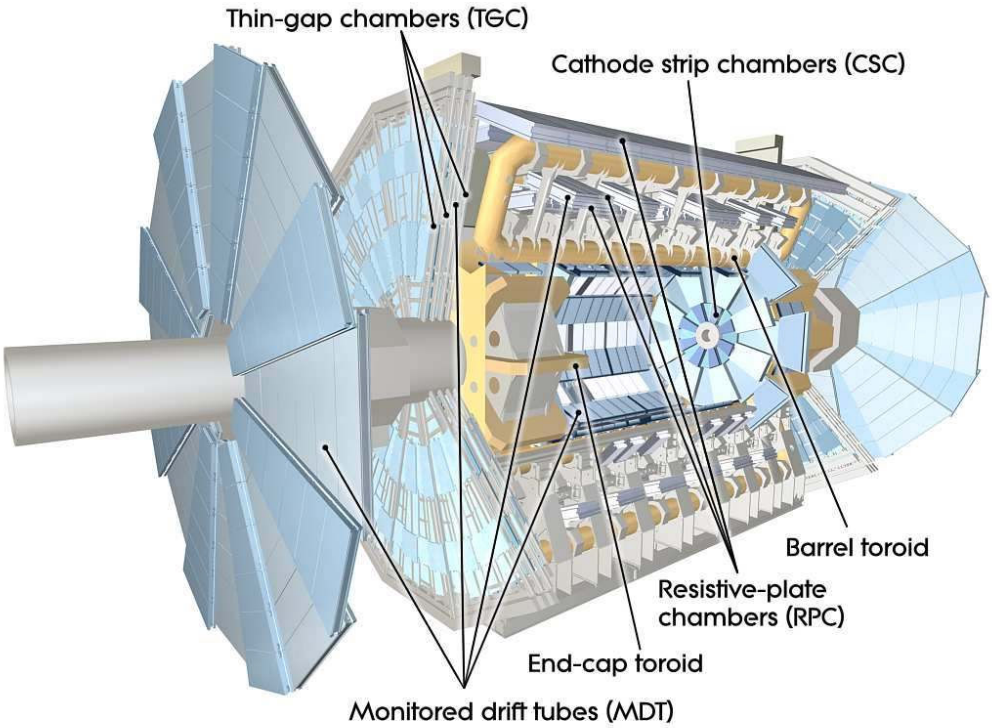


Figure 2.14.: Layout of the ATLAS detector muon systems.

2.2.3. Muon Systems

The Monitored Drift Tubes (MDT), Cathode Strip Chambers (CSC), Thin Gap Chambers (TGC) and Resistive Plate Chambers (RPC), together with the Toroid Magnets, form the muon spectrometer of the ATLAS detector. The layout of the muon systems can be seen in figure 2.14. They are instrumented with separate trigger systems and high-precision tracking chambers. They are an essential part of the detector; since muons are used in L1 trigger decisions; and they are important for studying many physics processes. The Muon spectrometer can independently measure the momentum of the muons from ~ 3 GeV to 1 TeV with varying resolutions.

The large superconducting air-core toroid magnets of the detector provide the magnetic field for deflecting the muon tracks. The barrel toroid magnets generate the magnetic field in the region $|\eta| < 1.4$ and the endcap magnets generate the field within $1.6 < |\eta| < 2.7$. In the region $1.4 < |\eta| < 1.6$, which is called the transition region, magnetic deflection is due to a combination of the field from the barrel and endcap toroids. In the range $|\eta| < 1.4$ the bending power generated by the barrel toroids is 1.5–5.5 Tm. The bending power due to the endcap toroids ranges from 1–7.5 Tm in the interval $1.4 < |\eta| < 2.7$. Figure 2.15 shows the integrated magnetic field strength as a function of η. The drop in the magnetic field strength in the transition region can be seen.

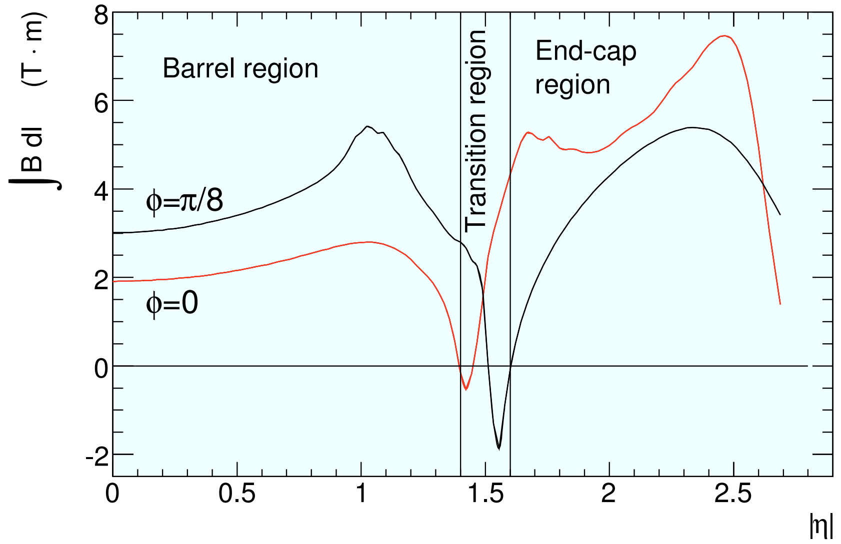


Figure 2.15.: η dependency of the integrated magnetic field strength for two different azimuthal angles [9].

The RPC and MDT chambers in the barrel region are arranged as three concentric rings, between and on the central toroid magnets, at radii of 5 m, 7.5 m and 10 m. In the endcap regions the muon chambers are in the form of large wheels, in front of and behind the endcap magnets, at locations $|z| \approx 7.4$ m, 10.8 m, 14 m and 21.5 m.

The RPC chambers are made up of two independent detector layers with gas volumes and two orthogonal sets of pickup strips. The RPC chambers in $|\eta| < 1.05$ and the TGC chambers in endcap regions within the range $1.05 < |\eta| < 2.4$, operate on the same principles as the multi-wire proportional chambers; providing fast track information. The timing resolution of the RPC and TGC chambers is about $15 - 25$ ns and both types of chambers can be triggered at high rates. The MDT chambers cover the ranges up to $|\eta| < 2.7$, providing about 35 μm resolution per module for the momentum measurements. In the region $2 < |\eta| < 2.7$, the CSC provides 40 μm resolution in the bending plane and about 5 mm in the transverse plane. Figure 2.16 shows the number of CSC and MDT stations traversed by muons at a given η and ϕ.

As described in [9] there are two different sets of algorithms called *Staco* and *Muid* for identifying and reconstructing the muons. In each algorithm set there are three different algorithms working either in the *standalone* mode; that is by finding the tracks in the muon spectrometer and extrapolating to the beamline or in *Combined* mode; by matching the standalone muons to the Inner Detector tracks and combining the measurements, or in *Tagged* mode; by extrapolating the Inner Detector tracks to the muon spectrometer for finding the muons. At design luminosity, the pileup events, low-energy photons and neutrons from the cavern background all affect the muon reconstruction efficiencies. These effects are estimated by monte-carlo simulations of $t\bar{t}$ events at a luminosity of $\mathcal{L} = 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The events are either used alone or overlaid with the

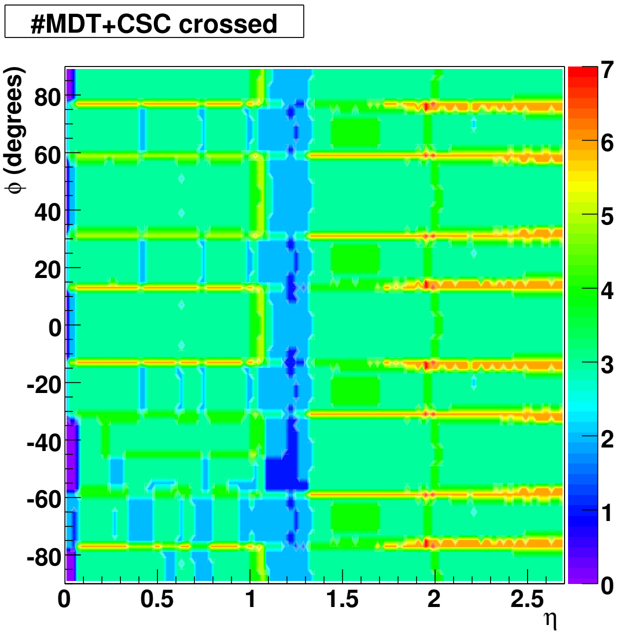


Figure 2.16.: Number of detector stations traversed by muons passing through the muon spectrometer as a function of ϕ and η. Taken from [9].

pileup and twice the expected cavern background. Those events overlaid with pileup are called "high luminosity sample" events and the others are called "low luminosity sample" events.

Figure 2.17 shows the reconstruction efficiencies and the fake rates for the direct muons in the $t\bar{t}$ events at two different luminosites for the Staco standalone algrithm *Muonboy* and the Muid algorithm *Moore/Muid*. The "good" efficiency muons require additional criteria to be satisfied; as described in [9]. Fake rates for different p_{T} thresholds are included in the plots in the second and fourth rows. The same distributions for the combined track algorithms are shown in figure 2.18.

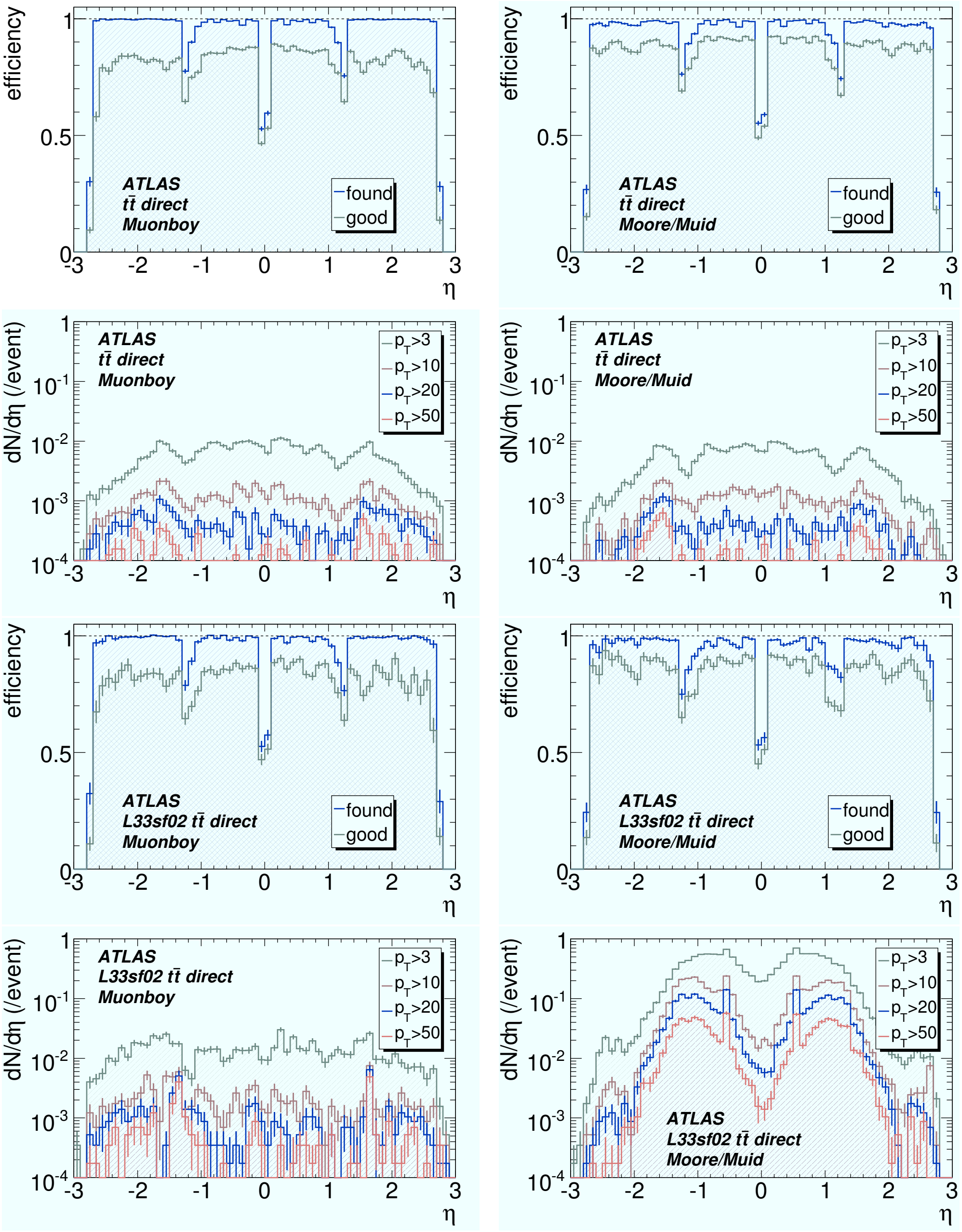


Figure 2.17.: Standalone efficiency and the fake rates for Staco and Muid algorithms for the direct muons from the $t\bar{t}$ events. The plots on the left are for the efficiencies and the fake rates of the Muonboy algorithm. The plots on the right are for the Moore/Muid. The top two rows are for the low luminosity (without pileup and cavern background) estimations and the bottom two are for the high luminosity (with pileup and cavern background). The second and fourth rows show the fake rates for different p_T thresholds [9].

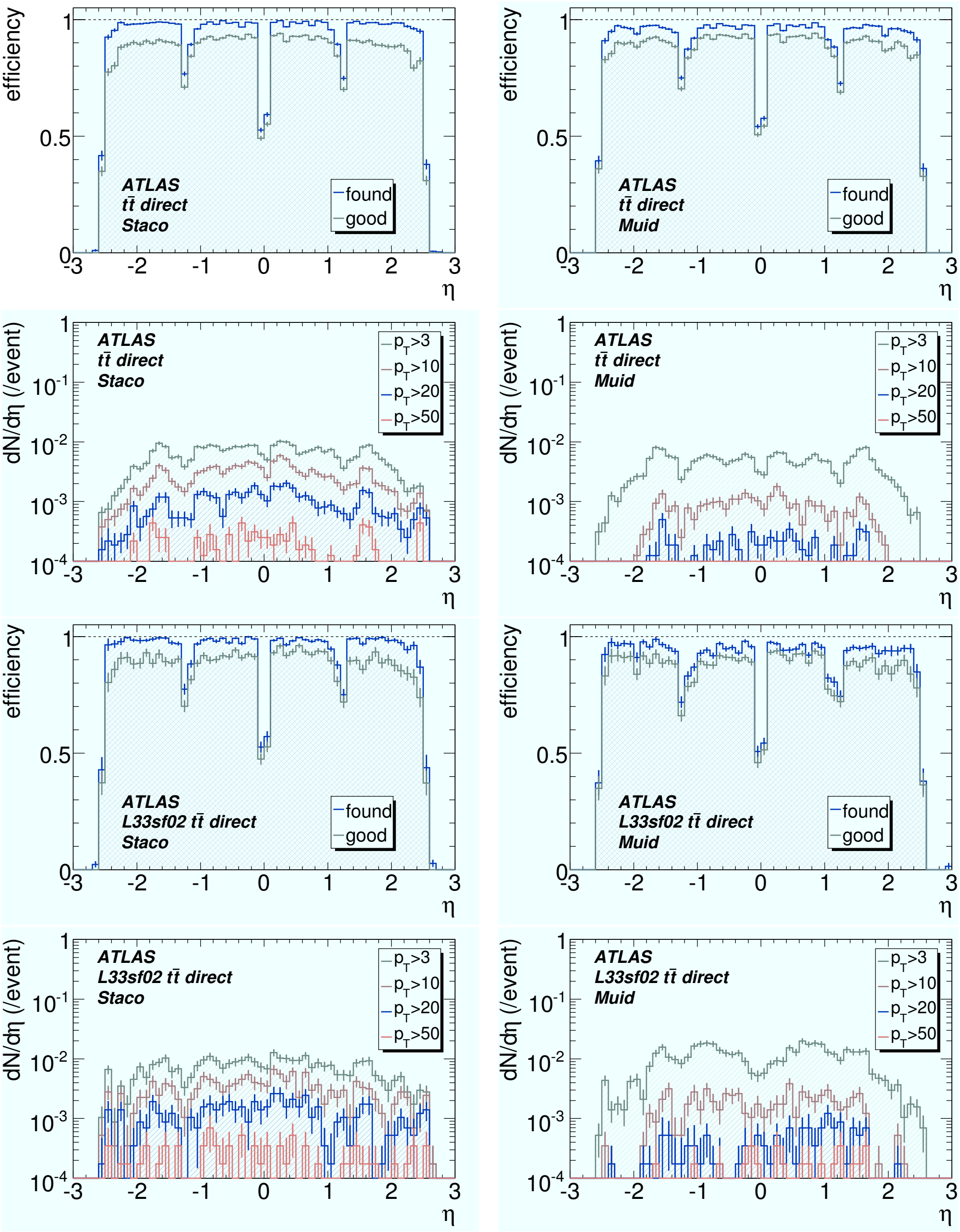


Figure 2.18.: Combined muon efficiency and fake rates for both algorithms for the direct muons from the $t\bar{t}$ events. The plots on the left are for the efficiencies and the fake rates of the Staco algorithm while the plots on the right are for the Muid. Top two rows are for the low luminosity (without pileup and cavern background) estimations and the bottom two are for high luminosity (with pileup and cavern background). The second and fourth rows show the fake rates for different p_T thresholds [9].

3. Trigger and Data Acquisition in ATLAS

About 18 interactions per bunch-crossing are expected during the LHC runs at 14 TeV and $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [12]. At 40 MHz bunch-crossing rate, this corresponds to ~ 720 million interactions per second. Such an enormous number of interactions require a quick and highly selective trigger. To accomplish this, ATLAS uses a three level triggering system. The "Level-1" (L1) trigger is implemented in hardware and reduces the event rate to 75 kHz. The "Level-2" (L2) trigger is implemented in software using commodity PCs and further reduces the event rate down to 3.5 kHz. Finally the "Event Filter" (EF) does full reconstruction and selects events to be recorded on disks, reducing the event rate down to 200 Hz. L2 and EF are together called the "High Level Trigger" (HLT). A block diagram of the trigger systems is shown in figure 3.1 and an overview of each level is presented in the following sections. Details of the ATLAS trigger and data acquisition (TDAQ) systems are available elsewhere [4, 13].

3.1. Level-1 Trigger

The Level-1 Trigger is responsible for reducing the event rate to a manageable level while keeping interesting events. It is implemented in hardware and reduces the event rate down to 75 kHz. This rate is limited by the data rate that the detector readout systems can handle, but it is possible to upgrade the L1 trigger to 100 kHz. The L1 trigger decision must reach the front-end electronics within $2.5\,\mu$s of the bunch-crossing that it is associated with. About $1\,\mu$s of this time is spent in signal propagation in cables. Because of these constraints the L1 uses only part of the information available with reduced granularity. It uses the information from the Resistive Plate Chambers (RPC) and Thin-Gap Chambers (TGC) to trigger on high p_{T} muons and information from the calorimeters to trigger on $E_{\mathrm{T}}^{\mathrm{miss}}$, τ-lepton, jets, EM clusters and large transverse energy. A block diagram of the L1 trigger is given in figure 3.2. The calorimeter information is processed by the Level-1 Calorimeter (L1Calo) trigger module and the muon information is handled by the Level-1 Muon trigger module. Each module reads the information from their respective detectors and passes the decisions to the Central Trigger Processor (CTP).

The CTP combines information from the muon and calorimeter modules for different trigger types and makes the final decision. It can be programmed for up to 256 sets of requirement combinations on the input data called trigger menus. The "Timing, Trigger and Control" (TTC) system delivers the decision, together with the 40.08 MHz clock and other signals to the front-end electronics through an optical-broadcast network. Although the CTP reports only the flags whose thresholds were passed, it keeps the

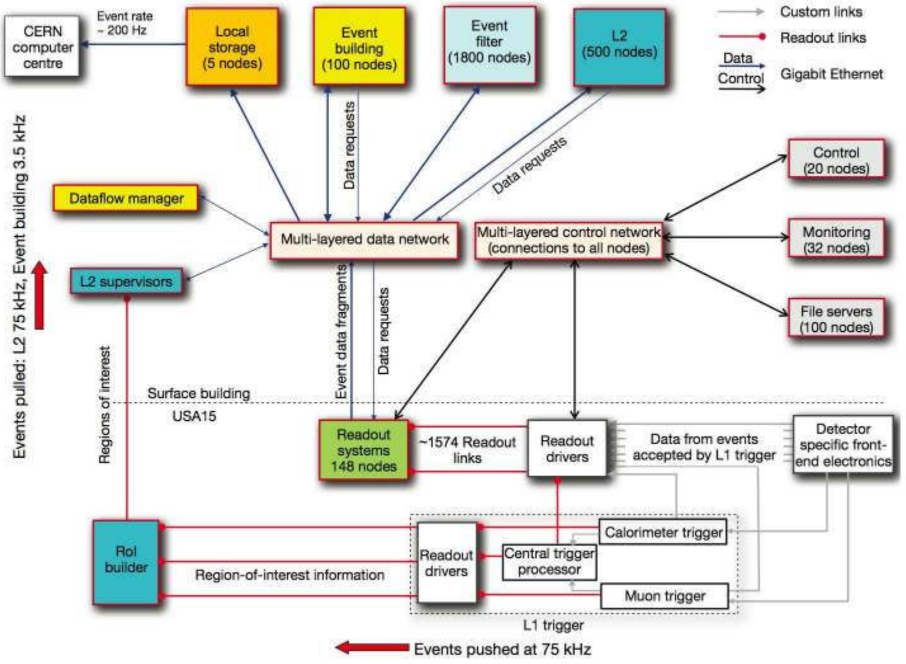


Figure 3.1.: A block diagram of the ATLAS Trigger, showing the data flow between the trigger elements. A basic Level-1 trigger diagram is shown inside the dash-lined box at the bottom of the diagram. Elements below the dashed line are located in the underground cavern (USA15) and elements above it are located in the surface building.

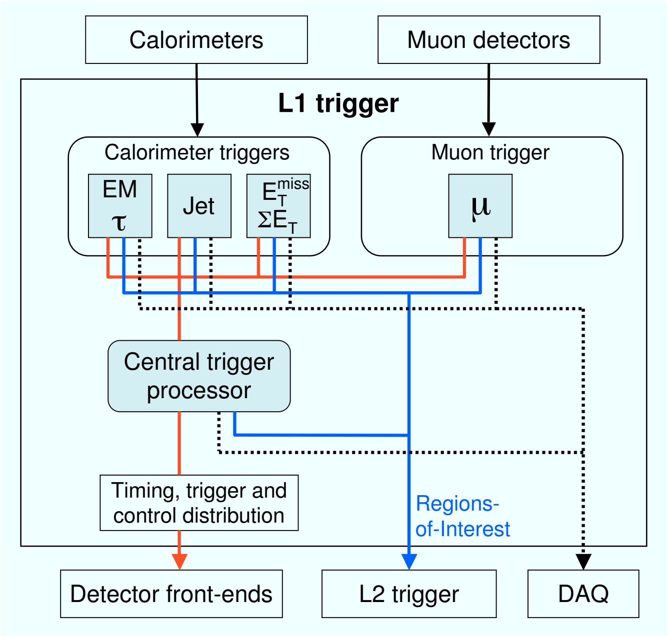


Figure 3.2.: Block diagram of the Level-1 trigger.

geometrical location of the trigger objects, called Regions of Interest (RoI), for seeding the HLT.

Level-1 Calorimeter Trigger

The Level-1 calorimeter trigger (L1Calo) uses the information from all calorimeters to identify the objects with high-E_T such as photons, electrons, τ-leptons decaying into hadrons and high-E_T^{miss} or very high $\sum E_T$.

Approximately 7000 analogue links made of twisted pair cables carry the reduced granularity information about the trigger towers from the EM and hadronic calorimeters to the analogue front-end electronics located at the USA15 cavern near the detector. There the signals are converted to digital and then go through bunch-crossing identification and energy estimation by a table look-up. The signals are further processed before they are sent to the Jet/Energy-Sum Processor (JEP) or Cluster Processor Module (CPM). Figure 3.3 shows a data flow diagram for the L1Calo triggers.

The Cluster Processor Module (CPM) searches for electrons/photons or τ-jets in a sliding window of 2 x 2 towers inside 4 x 4 tower regions. The electron/photon cluster trigger requires the sum of any row or column in the 2 x 2 window to be above a predefined threshold. The τ-trigger also uses the same method but includes the hadronic towers behind the the EM towers in calculations. It is also possible to require an isolation ring, the 12-towers surrounding 2 x 2 region in question, to be below a predefined threshold. Figure 3.4 shows a diagram of the region and the surrounding ring. Since several neighboring 2 x 2 windows can satisfy the trigger conditions, a special care is

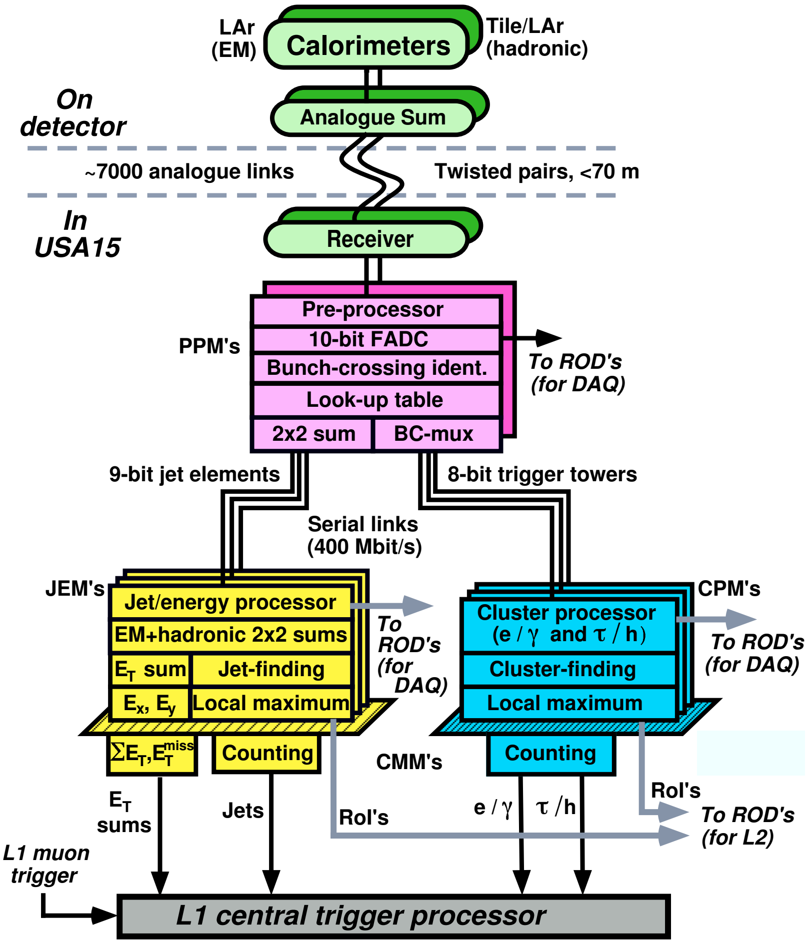


Figure 3.3.: Level-1 calorimeter trigger data flow diagram. The Preprocessor module block is in purple, the Jet Energy Module (JEM) is in yellow and the Cluster Processor Module (CPM) is in turquoise.

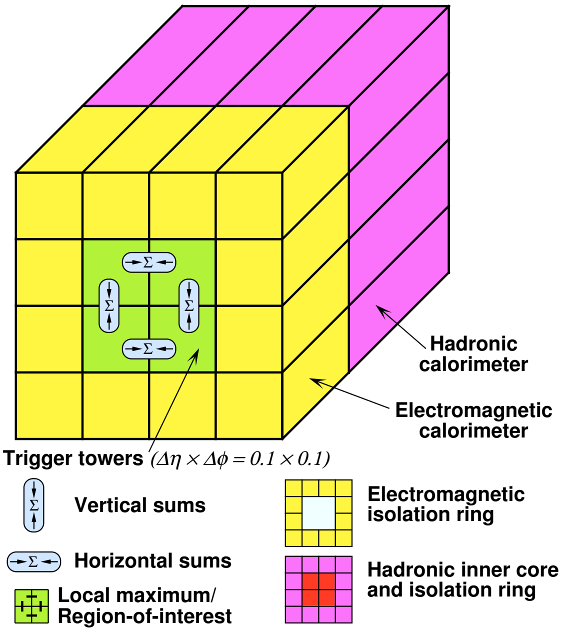


Figure 3.4.: Schema of the trigger algorithm used in the Cluster Processor Module.

taken to avoid multiple counting. The local maximum in such cases is defined as the region where four of the eight possible neighboring windows have equal or less energy deposition while the other four have less energy deposition than the selected region. The local maximum also defines the RoI's for the electron/photon and τ-triggers.

The Jet/Energy Module (JEM) works with the jet elements which are the sums of 2 x 2 towers in the EM calorimeters and the hadronic calorimeters behind them. The JEM calculates E_T on overlapping windows of sizes 2×2, 3×3 or 4×4 jet elements and compares them to predefined thresholds. Similar to the electron/photon and τ-triggers, multiple counting is avoided by requiring the 2 x 2 jet element window to be a local maximum. The same window is defined as the jet RoI. The JEM can compare the results with four $\sum E_\mathrm{T}$ and eight $E_\mathrm{T}^\mathrm{miss}$ thresholds and report results to the CTP.

3.1.1. Level-1 Muon Trigger

The L1 Muon trigger uses the information from the RPC's in the barrel region and the TGC's in the end-cap region. The fast response from the RPC's and TGC's enables L1 Muon trigger to associate the muon tracks with their corresponding beam-crossings. The L1 Muon trigger checks for muon tracks by looking at the coincidence hits in different chamber layers within a certain path defined by the trigger p_T thresholds. Both the end-cap and barrel results are combined together to form six threshold multiplicities, three of them for low-p_T and three for high-p_T thresholds, before they are passed to the CTP.

The L1 Muon barrel trigger tries to find the muons over the thresholds by using the hits in the barrel RPC layers. It starts from the hits in the middle RPC layer and looks for a matching hit in the inner layer within a region which is centered on the line connecting the middle layer hit and the interaction point. The width of the region depends on the p_{T} threshold. Requiring at least 3 hits out 4 possible in the middle and inner RPC layers together reduces the fake tracks from noise hits, accidental triggers from low-energy particles in the cavern and increases the trigger stability.

For high-p_{T} triggers, low-p_{T} trigger results are checked with higher thresholds and for the matching hits on the outermost RPC layer in a similar way that is done for the low-p_{T} trigger. Like the low-p_{T} trigger, at least one hit requirement in the outermost RPC layer together with the low-p_{T} requirement increases the stability while reducing the fake triggers. The low-p_{T} and high-p_{T} triggers both have three trigger thresholds within the ranges 6–9 GeV and 9–35 GeV respectively. All six thresholds are simultaneously used in the trigger. These results together with the η and ϕ information form the RoI for the L1 Muon trigger.

The L1 Muon end-cap trigger uses the hits in the TGC and operates in a similar way to the barrel trigger. It searches for the matching hits in the first and last TGC plane by looking at the hits within the region centered on the line connecting the hit in the middle TGC plane to the interaction point. Like the barrel trigger, all thresholds are simultaneously checked and the results are converted to RoI's by combining with the coordinate information. The results from the barrel and end-cap regions are combined and total multiplicities per thresholds are sent to the CTP.

3.1.2. Central Trigger Processor

The Central Trigger Processor receives the trigger information from the L1Calo and the L1Muon triggers and additional information from the beam-pickup monitors for the filled-bunch trigger, and from the scintillation counters for the minimum-bias trigger (MBTS). Received trigger informations are converted to trigger conditions through look-up-tables. The CTP can support up to 256 trigger conditions including the trigger conditions derived from two random triggers, two pre-scaled clocks and eight triggers for programmable groups of bunch crossings. The trigger conditions are combined to form up to 256 trigger items similar to *at least one muon and two or more jets above given thresholds*. Then the CTP generates a Level-1 Accept (L1A) signal by logically OR'ing all available trigger items. When the L1A is true, the trigger decisions for all the trigger items are sent to the Level-2 trigger through the RoI Builder (RoIB) and Read Out Systems (ROS).

Read Out Systems receive the event fragments through 1574 Read Out Links (ROL) from the detector-specific Read-Out Drivers (ROD) into the Read-Out Buffers (ROB) they contain. The RoIB receives the RoI information from the L1 trigger modules. Then it assembles these information into a single RoI data structure and transfers this structure to Level-2 SuperVisors (L2SV) using a round-robin algorithm.

The CTP tags the data with a Luminosity Block (LB) number. A luminosity block is defined as the shortest time interval where the corrected integrated luminosity can be determined. At each LB change, the CTP values such as the number of triggers generated by the trigger logic, surviving the pre-scale and dead-time vetoes are stored together with the LB in a database for offline determination of the luminosity corrections. The CTP also provides finer detailed values for the monitoring purposes. Additionally it manages the clock signal of the detector, distributing it to all other sub-systems synchronized with the LHC clock.

3.2. Data Acquisition and High-Level Trigger

The L1A signal triggers the transfer of the data in the front-end electronics to the Read-Out Drivers (ROS) and from there to the Read-Out Systems (ROS). Meanwhile the RoI Builder (RoIB) combines the RoI produced by the L1 trigger modules creating a RoI data structure. The processing of the event in the HLT begins with the arrival of the RoI structure to the Level-2 SuperVisor (L2SV). Figure 3.5 shows the event flow in the HLT. The L2SV supervise the flow of events at Level-2. L2SV assigns the RoI from L1 to one of the Level-2 Processor Units (L2PU). Using the information in the RoI, the L2PU asks for the respective event fragment from the corresponding ROS. After it receives the fragment, information is processed and the results are sent back to the L2SV. The L2SV forwards the result to the Data Flow Manager (DFM). If the event is not accepted in Level-2, the DFM tells all the ROS's to discard the respective event data from their buffers. Otherwise it assigns the event to a Sub-Farm Input (SFI) node. The SFI collects the event fragments from the ROS's and builds the event-data structure. Upon successful assembly of the event-data structure, the SFI notifies the DFM so that it can order the ROS's to dump the respective event data from their buffers. The completed event-data is sent to an Event-Filter Data-flow (EFD) node. The EFD assigns the event to one of the Processing Tasks (PT) in the event filter farms. In addition to event selection, the PT classifies the event into a set of predefined event streams. The results from the PT are sent back to the EFD where the selected events are forwarded to the Sub-Farm Output (SFO) nodes. The SFO writes the event to the respective event stream or streams and sends them to CERN's central data recording facility.

The DAQ/HLT system is composed of many computers and a large number of software components working in a distributed manner. For this reason the DAQ/HLT control system is also implemented in a distributed manner. Two main components of the control system are the process management daemon and the Run Controllers.

The Process management daemon runs on every computer. It handles the execution and termination of processes with the settings described in the configuration databases, taking resource availability and execution permissions into account. The Run Controllers on the other hand steer the data acquisition by sending necessary commands to the processes like start, stop or change state in the finite state machine. Run Controllers follow a tree structure which is related to the detector system and sub-system decomposition. The error and information messages generated by all processes including the control systems are archived in a database via the Log Service.

The PC's used in the HLT reside behind an application gateway which enables access to the experiment's local area network using a role-based access control scheme. All PC's

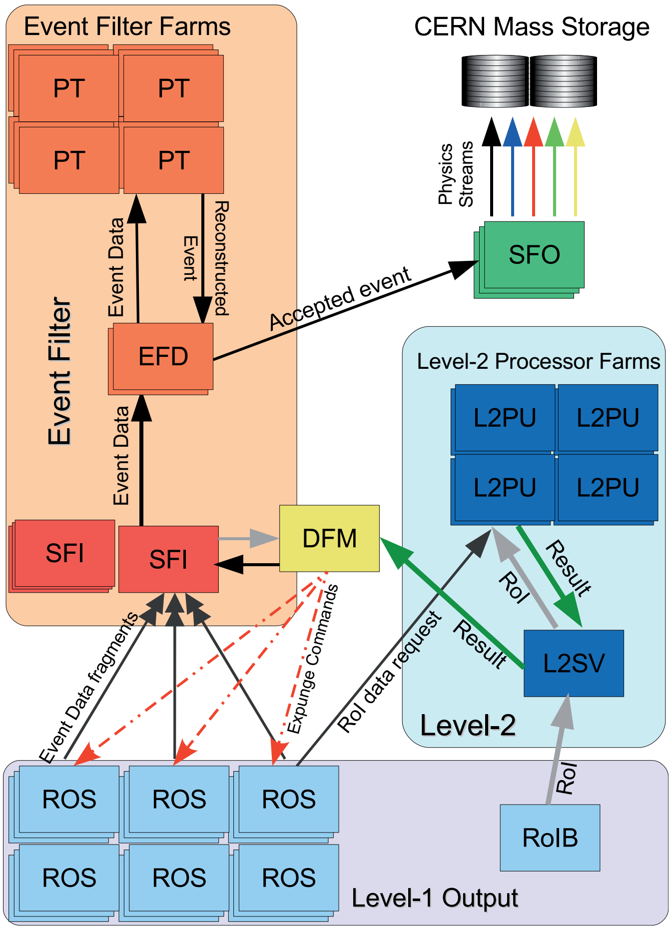


Figure 3.5.: Event flow of an accepted event in High Level Trigger from Level-1 output to the CERN data-recording facility as described in section 3.2.

are booted to a reduced version of Scientific Linux CERN (SLC) through network based boot images located in the Local File Servers (LFS). Anything that is not essential for the operation of the PC's are excluded from the boot images and are available through the networked file system, together with the ATLAS software, from the LFS. A Central File Server distributes the master image of the software to the LFS's providing uniformity over the computing cluster. The health of the computers are monitored through customized software which can send SMS and e-mail notifications to the responsibles.

The configuration information of the hardware and software components of the trigger systems are kept in configuration databases, in the form of linked segments following the detector and DAQ/HLT hierarchy. In order to be able to cope with the requests from thousands of processes the databases are replicated in several servers.

The configuration of the thresholds for triggering at L1 and the trigger algorithms at L2 and EF are kept coherent and consistent with the help of an integrated system which stores the necessary values in a relational database. This database is accessible by all nodes in the HLT and the systems taking part in the L1 trigger decision.

The tree structure in the configuration and control systems makes it possible to run parts of the detector independently. Such parts are called partitions. Once in operation the partitions can be combined and run as a single partition up to a fully integrated and operational detector.

For reasons of cost, initial deployment of the DAQ/HLT farms is only a subset of what is foreseen. Table 3.1 lists the numbers of initially deployed PC's, their CPU types, memory size and number of CPUs per node. The PC's, except the ROS's, are placed in the standard 47U or 52U sized server racks. These racks are cooled by the water cooled heat-exchangers and the horizontal airflow within the rack. Each rack is equipped with a LSF and two gigabit-Ethernet switches and typically contain about 30 1U PC's. Assuming 1.3 MB average event size, approximately 2.0 kHz of event-building rate can be sustained by the 48 SFI deployed for the initial operations.

The SFO PC's contain three RAID controllers, each connected to eight 500 GB SATA-II disks. These disks provide enough buffer for about 24 hours in case connection to CERN data-storage is interrupted for some reason. The SFO writes the event data to these disks with respect to their physics streams. In order to maximize the disk throughput, the controllers are used in the form of a circular buffer such that when the data in a set of disks are read for transferring to the CERN data-storage, the events are written into another set. In this manner a controller is used either for the write or read operations. Thus, with the six deployed SFO's, it is possible to have a sustained output of 300 MB/s and a peak rate of 600 MB/s which corresponds to about 200 Hz event rate for a 1.3 MB average event size.

During the design phase of the DAQ/HLT, the CPU clock speeds were estimated to be in the order of 8 GHz, where processing of an event would take 10 ms in the L2PU and 1 s in the EF. This prediction turned out to be incorrect, however nowadays many CPU cores, running at slower clock speeds, are placed in one package. Thus a 2.0 GHz quad-core CPU running 4 processes in parallel would process 4 events in $\sim$ 40 ms at L2PU and in $\sim$ 4 s at EF, effectively satisfying the design criteria.

Due to the parallel structure, the ATLAS trigger has some extend of redundancy and

Component	**Number of nodes**	**Number of racks**	**CPU's per node**	**Memory (GB)**	**CPU type**
ROS	145	16	1	0.5	3.4 GHz Irwindale
SFI	48	3	2	2	2.6 GHz Opteron 252
DFM	12	1			
L2SV	10				
HLT	1116	36	8	8	1.86 GHz Xeon E5320
SFO	6	2	2	4	2.00 GHz Xeon E5130
Monitoring	32	4	4	8	3.00 GHz Xeon E5160
Operations	20		2	4	2.00 GHz Xeon E5130

Table 3.1.: Number of PC's deployed for the initial operation of the detector. During initial operation phase HLT nodes are assigned to L2 and EF farms depending on the run conditions.

can continue working with a reduced effective rate when one or more L2 or EF processes are terminated.

Depending on the data taking conditions, initial set of HLT racks will be distributed between the EF and L2 farms and it is estimated that the high level trigger can handle about 40 kHz L1 trigger rate which corresponds to about half of the design rate [4].

The L2 and EF processes use an early rejection approach to satisfy the required event selection rates. A trigger decision is divided into several steps called trigger chains. Each step is seeded by the output of the previous step and produces the result by running a quick algorithm. If the result do not satisfy the requirements, the chain fails. If all chains fail for an event, it is discarded. Figure 3.6 shows an example of a L2 trigger chain, "e25i", requiring an isolated electron shower with an energy greater than 25 GeV. The chain starts with L1 output, an activity in the EM calorimeter satisfying the L1 isolated EM calorimeter hits with an energy of about 25 GeV (EM25I) trigger. Then it applies a shower shape algorithm to see whether it looks like an electron shower, producing the next step, an electron candidate. The track algorithm checks for a track pointing towards the shower, approving the shower as an electron shower. In the next step, the energy algorithm calculates the energy of the shower comparing it to the 25 GeV threshold. Finally the isolation condition is tested by the isolation algorithm. If the event satisfies all of the conditions, it is sent to the EF for further checks. The information produced by the L2 algorithms are sent to another application, pseudo-ROS, to be included in the event structure by the SFI. A similar triggering schema, seeded by the L2 output is applied at the EF farms. Results of the EF trigger chains determine the physics stream or streams where the event will be recorded.

The combination of all the trigger items and chains in a run is called a Trigger Menu. It defines the thresholds and the algorithms used in all trigger levels. The combinations of the trigger chains are defined by the physics goals. Figure 3.7 shows the expected

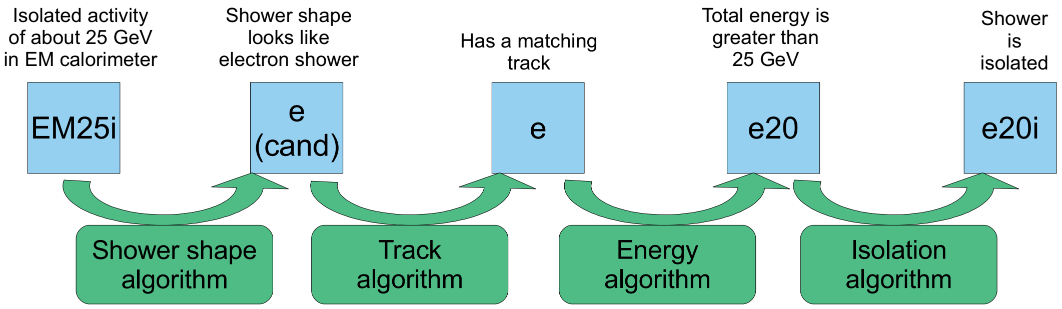


Figure 3.6.: An example trigger chain for L2, requiring an isolated electron shower with energy greater than 25 GeV in the EM calorimeter.

Abbreviation		**Threshold Type**
L1	L2 and EF	
EM	em	Electromagnetic
J	j	Jet
TAU	tau	Tau
JF	fj	Forward Jets
TE	te	Total transverse energy
XE	xe	Missing Energy
JE	je	Total Energy from Jets
MU	mu	muon
-	mb	minimum bias
-	e	electron
-	g	gamma

Table 3.2.: Trigger type abbreviations.

event rates and cross-sections with respect to jet E_T or particle mass [9].

The names of the trigger items follow the pattern

[Level]_[Multiplier][Type][Threshold(GeV)][Add. Info]

such as *EF_2mu20*, meaning at least 2 muons with $p_T > 20$ at EF or *L1_TE650*, total scalar transverse energy $\sum E_T > 650$ at L1. Combined trigger items are separated by an "_" and logically "AND"ed. The *Type* are abbreviations, listed in table 3.2. All L1 items are in capitals and L2 and EF items are in lower case. Intermediate trigger elements in L2 and EF have an additional field which describes the algorithm.

L1 trigger items are pre-scaled, i.e. only 1 in N events are used, in order to keep their rates under control. Table 3.3 contains a list of L1 trigger items, pre-scales and expected rates for a luminosity of $\mathcal{L} = 10^{31}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ and figure 3.8 shows the expected jet and muon rates at the L1.

The inclusive jet triggers with low E_T are highly pre-scaled in order to keep the L1 bandwidth small. For the low-luminosity run, inclusive jet triggers are used in pass-through mode in HLT, i.e accepted independent of the L2 and EF trigger decisions. In this way enough data can be collected both for differential cross-section measurement and validation of the algorithms in the HLT. On the other hand multi-jet triggers have smaller pre-scale values, thus higher bandwidth, to allow the selection algorithms which can only run in HLT, such as b-tagging algorithms, to be used.

In contrast to inclusive jet triggers, single muon and di-muon trigger item rates are not pre-scaled in L1. The rates are controlled by the HLT selection. The tau triggers are selected for collecting a large number of W and Z samples where they decay to τ leptons. Since the $E_\mathrm{T}^\mathrm{miss}$ triggers are highly sensitive to the detector efficiency and acceptance effects in the beginning, they are expected to be less efficient and used in conjunction with other trigger items. Similarly $\sum E_\mathrm{T}$ and the total energy from the jets trigger items will be operating with lower efficiency until the detector effects are understood. Multi-trigger items are deployed together with the single trigger items in order to validate them for high-luminosity runs.

Most of the L1 triggers are configured in pass-through mode in the HLT for the low-luminosity run. The lowest-threshold HLT trigger items and their expected rates at $\mathcal{L} = 10^{31}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ are given in table 3.4.

The details of triggering procedure in the HLT and detailed explanations of the trigger chains are available in [9, 13] and the references therein.

3.3. Monitoring of ATLAS Trigger and DAQ

In a distributed and complicated system such as ATLAS DAQ/HLT, many things can go wrong, some interrupting the operation, some going unnoticed. In order to prevent failures and create a running condition as stable as possible, the DAQ/HLT systems should be monitored at all times. Monitoring of the DAQ/HLT is achieved through the “Information Sharing Services”. There are three different types of services for different classes of information, the “Information Service” (IS) for basic variable types, the “On-line Histogramming” (OH) for publishing histograms and the “Event Monitoring” (Emon) for event data monitoring.

The Information Service provides the distribution of the values of the structures made up of basic variable types. It provides a subscription mechanism for the information consumers to get notifications of the changes to one or more items. It also enables any application to send commands to any information provider. Almost all applications in DAQ/HLT publish information about themselves and other quantities relevant with their tasks to the IS servers in one or more structures of basic variables called “IS Objects”. There are approximately 100 IS servers running in the initial DAQ/HLT setup.

The On-line Histogramming service is an extension of the IS to support raw and ROOT histograms. L2PU and EF applications produce histograms about the events that are processed. These histograms are published to OH servers on certain time intervals. The

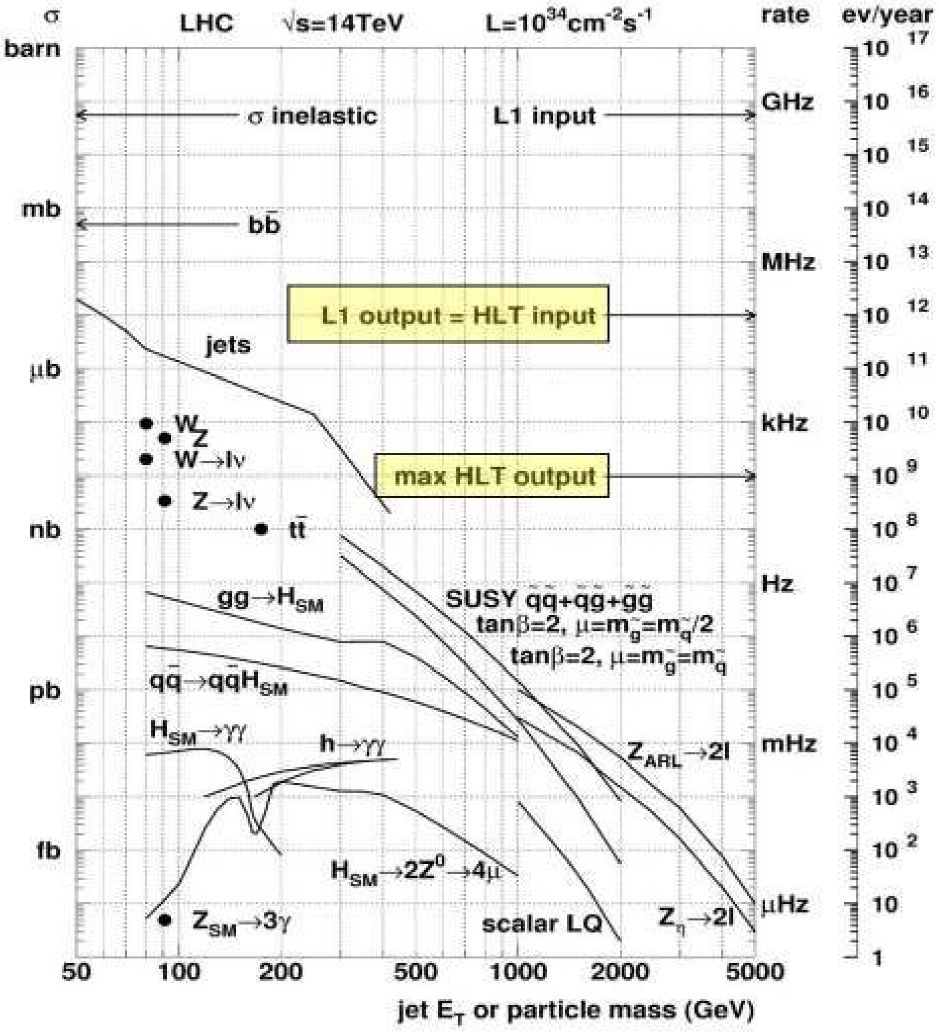


Figure 3.7.: Expected event rates at the LHC design luminosity [9].

Trigger Item	**Pre-scale**	**Rate (Hz)**	**Trigger Item**	**Pre-scale**	**Rate (Hz)**
Electromagnetic Objects (EM)					
EM3	60	674	EM7	1	4900
EM13	1	950	EM13I	1	480
EM18	1	369	EM18I	1	143
EM23I	1	53	EM100	1	15
2EM3	1	6500	2EM7	1	534
2EM13	1	108	2EM13I	1	8
2EM18	1	47	2EM18I	1	2
2EM23I	1	0.6	3EM7	1	53
Jet Objects (J)					
J10	42000	4	J18	6000	1
J23	2000	1	J35	500	1
J42	100	4	J70	15	4
J120	1	9	3J10	150	40
3J18	1	140	4J10	30	40
4J18	1	20	4J23	1	7
Forward Jet Objects (FJ)					
FJ18	7000	1	FJ35	700	1
FJ70	20	1	FJ120	1	1
2FJ18	100	1	2FJ35	1	2
Muon Objects (MU)					
MU4	1	1730	MU6	1	640
MU10	1	360	MU15	1	30
MU20	1	20	MU40	1	10
2MU4	1	70	MU4_MU6	1	45
2MU6	1	14	2MU10	1	7
2MU20	1	0.2	3MU6	1	0.7
τ Objects (TAU)					
TAU6	750	19	TAU9I	300	16
TAU11I	1500	2	TAU16I	10000	<0.1
TAU25	20	16.1	TAU25I	10	25
TAU40	1	83	2TAU6	100	19
2TAU9I	1	413	2TAU16I	1	65
TAU6_TAU16I	10	46	TAU9I_EM13I	1	100
TAU9I_MU6	1	25	TAU9I_XE30	1	160
Missing E_{T} Objects (XE)					
XE15	30000	2.5	XE200	7000	3
XE25	1500	4	XE30	200	7.5
XE40	20	7.5	XE50	2	14
XE70	1	2	XE80	1	1
Scalar $\sum E_{\mathrm{T}}$(TE) and $\sum E_{\mathrm{T}}^{\mathrm{jet}}$ (JE) Objects					
TE150	100k	2	TE250	1100	3
TE360	40	1	TE650	1	0.5
JE120	150	0.5	JE220	10	0.5
JE280	2	0.5	JE340	1	0.1
Combination Objects (A and B)					
EM13_XE20	1	225	EM7_MU6	1	10
MU11_XE15	1	13	MU10_J18	1	33
2J42_XE30	1	13	4J23_EM13I	1	6.5

Table 3.3.: Some of the foreseen L1 trigger items for $\mathcal{L} = 10^{31}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

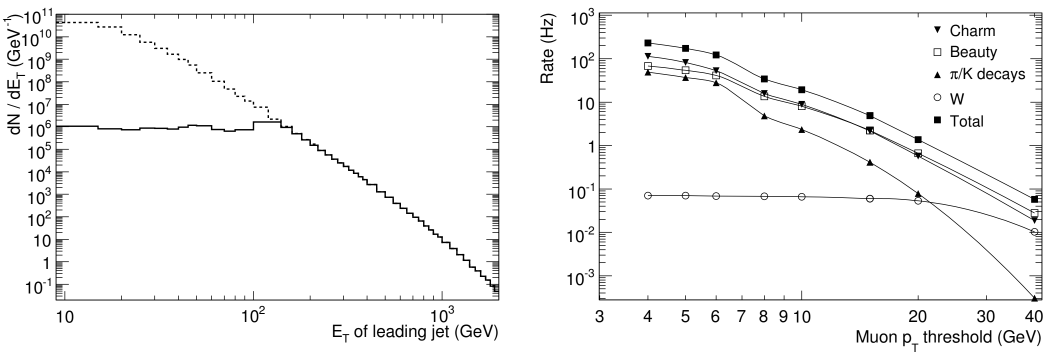


Figure 3.8.: Expected event rates at L1. The left plot shows the jet-E_T spectrum with (solid) and without (dashed) pre-scaling. The right plot shows the muon trigger rates at different thresholds, together with the contributions from various processes.

Trigger Item	**Rate**	**Trigger Item**	**Rate**
e12	19	2e5	7
g20	7	tau60	10
tau25i_XE30	3.5	MU10	18
2MU4	2.3	e10_MU6	0.5
J120	9	4J23	7
2b23	3		

Table 3.4.: Some L2 and EF trigger items and their expected rates foreseen for $\mathcal{L} = 10^{31}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

Gatherer application adds up these histograms, publishing the results back. The OH servers follow a tree structure. Usually there is one OH server per PC rack in the HLT and a corresponding instance of the Gatherer application, publishing results to the OH server on a higher level. The histograms in the OH servers are displayed in ROOT and Qt based applications which support overlaying reference histograms, sending commands to publishers and custom plug-ins.

The Data Quality Monitoring Framework (DQMF) is implemented in order to automatically analyze the large quantities of histograms produced by trigger processes. It can compare histograms with the user-provided reference histograms or values using the user-supplied algorithms and can generate alarms when deviations occur. The results of checks are published to a special DQM IS server. A shifter can monitor results of DQMF checks by using the "DQMDisplay" application.

The events are monitored through the Event Monitoring service. The Emon service can read event fragments from the ROS's or full event data from SFIs, depending on the event type, and can ask for reconstruction of events in specially configured EF applications. The results of reconstruction can be used in applications visualizing the event data.

An extensive amount of information is collected in Information Sharing Services, however, this information must be presented to the shifters and experts in a compact, yet representative way. There are different tools prepared for that, each for a different purpose. Figure 3.9 shows a diagram of information providers and main programs used for displaying the information to the shifters and experts.

The OH Display and the On-line Histogram Presenter (OHP) are used for displaying the histograms in the OH servers. The OH Display displays the user-selected histograms and is used for basic checks rather than monitoring. The OHP on the other hand can retrieve many histograms through subscription mechanism or request and display them together with the reference histograms when provided. Its user interface can be configured to display many histograms in tabs and provides a plug-in support where the user interface can be specialized for different purposes. One such plug-in is "TriP", where the trigger chain information published by the L2 and EF nodes are displayed at different detail levels together with the CPU and memory utilization of the nodes.

The IS Monitor and Operational Monitoring Display (OMD) are used for displaying the information in the IS servers. The IS Monitor displays the values of an IS Object in a tabular manner and is usually used by the experts. The OMD, on the other hand, is prepared keeping both shifters and experts in mind. It is a highly configurable tool which can display the values in the IS objects as time-series graphs, bar-charts, distributions or numbers in tables.

The OMD uses the subscription mechanism for retrieving the IS Objects from IS servers. It can classify the objects with the help of regular expressions and calculate the averages, standard deviations and sums of the objects. It is configured through the user interface, mainly with drag-and-drop actions. The configuration can be altered in real-time and different levels of detail are available with a couple of mouse clicks. Up to 10 plots can be displayed in $N \times M$ format in many tabs and each tab can be detached from the main window on user request.

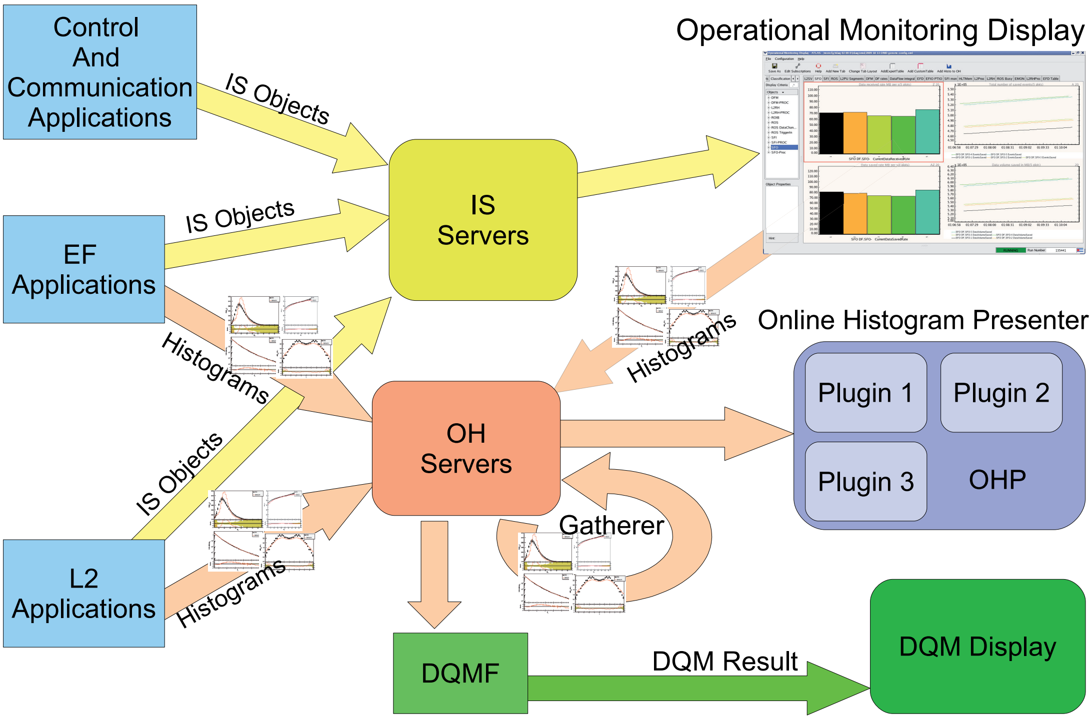


Figure 3.9.: A diagram showing the main monitoring displays and their information providers. The IS Objects from almost all applications are published to approximately 100 IS servers. The applications in the L2 and EF farms publish the information about the trigger chains and other observables in the form of histograms to OH servers. The Gatherer collects these histograms and publish the sums back. The DQMF checks the distributions in the histograms. The OHP and its plug-ins display histograms to shifters. The OMD retrieves the IS Objects, displays the information to the shifter and can publish the histograms of objects to the OH servers.

Figure 3.10.: A picture showing the main OMD window displaying different quantities in various types of plots. On the left panel the objects are listed in tree structures. On the right panel the distribution, bar-charts and time-series plots can be seen. The bar-charts can display either the instantaneous value or histogram.

The OMD can create one or two dimensional histograms out of the IS objects and publish them to OH servers, enabling the histogram processing applications to access the IS information easily. It can read multiple configuration files and create a combined display. It is also possible to run it in a GUI-less mode where it can work as part of a partition and provide the histograms of IS objects for other applications in the background. It also supports basic threshold checks on certain fields of the IS objects and can print a message when the value of the object is above, below, equal to or not equal to a given threshold. Figure 3.10 shows a snapshot of the main window with the different plot types available in OMD. Details of the program are given in appendix A.

4. Monte-Carlo Programs

4.1. The Standard Model

The Standard Model (SM) is a SU(3) × SU(2) × U(1) gauge field theory [14, 15]. It explains the interactions of matter with the electromagnetic, weak and strong force carriers. In the Standard Model there are six leptons, six quarks and four types of force carriers. The leptons and quarks are fermions and the force carriers are bosons. The electromagnetic force is carried by the photons, $W^{\pm}$ and Z^0 bosons carry the weak force and the strong force is mediated by the gluons. Figure 4.1 shows the elementary particles and the force carriers in the Standard Model. In the Standard Model, masses of the particles are explained by spontaneous symmetry breaking and the Higgs Mechanism [16]. The Higgs mechanism requires the existence of a new scalar particle called Higgs (H^0) boson. So far the Higgs boson has not been observed and it is one of the major reasons that the LHC has been built.

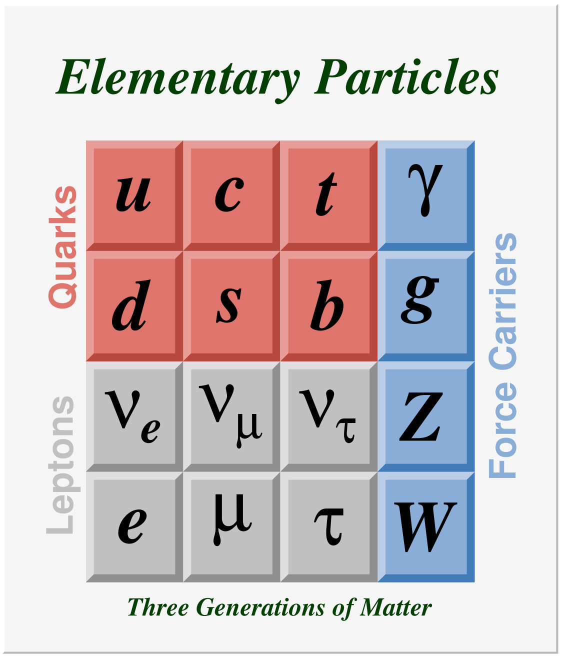


Figure 4.1.: Particles and force carriers in the Standard Model

Quarks and gluons are the constituents of the hadrons. The quarks carry fractional electric charge and a color charge and can interact electro-weakly or strongly. The gluons are electrically neutral and carry a pair of color charges, i.e. bicolored. They can interact only strongly and the interactions between the quarks and gluons are described by the Quantum Chromodynamics (QCD). It is a SU(3) symmetry with three colors, red, blue and green, thus quarks carry one color or anti-color on the other hand gluons carry a color and an anti-color charge. In QCD there are 8 gluons. The Lagrangian of the QCD can be split into a free-field part and an interaction part as

$$\mathcal{L}_{\mathrm{QCD}} = \mathcal{L}_{\mathrm{free}} + \mathcal{L}_{\mathrm{int}}.$$

Both terms can further split into quark, gluon and Faddeev-Popov (FP) ghost term, which is necessary for the elimination of unphysical particles from the theory [17]

$$\begin{aligned} \mathcal{L}_{\mathrm{free}} &= \mathcal{L}^{\mathrm{q}}_{\mathrm{free}} + \mathcal{L}^{\mathrm{g}}_{\mathrm{free}} + \mathcal{L}^{\mathrm{FP}}_{\mathrm{free}} \\ \mathcal{L}_{\mathrm{int}} &= \mathcal{L}^{\mathrm{qg}}_{\mathrm{int}} + \mathcal{L}^{\mathrm{gg}}_{\mathrm{int}} + \mathcal{L}^{\mathrm{FPg}}_{\mathrm{int}}. \end{aligned} \tag{4.1}$$

Free-field Lagrangian terms containing the kinetic terms of the quark, gluon and FP ghost fields can be written as

$$\mathcal{L}^{\mathrm{q}}_{\mathrm{free}} = i\sum_{j=1}^{n} \bar{\Psi}^{\alpha}_{j}\gamma^{\mu}(\partial_{\mu})_{\alpha\beta}\Psi^{\beta}_{j} - \sum_{j=1}^{n} m_j \bar{\Psi}^{\alpha}_{j}\Psi_{j,\alpha} \tag{4.2}$$

$$\mathcal{L}^{\mathrm{g}}_{\mathrm{free}} = -\frac{1}{4}(\partial^{\mu}A^{\nu}_{a} - \partial^{\nu}A^{\mu}_{a})(\partial^{\mu}A^{\nu}_{a} - \partial^{\nu}A^{\mu}_{a}) - \frac{1}{2\alpha_G}\partial^{\mu}A^{a}_{\mu}\partial_{\mu}A^{\mu}_{a} \tag{4.3}$$

$$\mathcal{L}^{\mathrm{FP}}_{\mathrm{free}} = -\partial_{\mu}\bar{\varphi}_{a}\partial^{\mu}\varphi^{a} \tag{4.4}$$

where A^{a}_{μ} are the gluon fields, Ψ_j and m_j are the field and the mass of the quark flavor j, respectively. $\varphi^{a}(x)$ are the eight anti-commuting scalar fields in SU(3) and the α_G is the gauge parameter. Similarly the interaction term can be written in three parts to describe the quark-gluon, gluon-gluon and FP-gluon interactions as

$$\mathcal{L}^{\mathrm{qg}}_{\mathrm{int}} = gA^{\mu}_{a}\sum_{j=1}^{n}\bar{\Psi}^{\alpha}_{j}\gamma^{\mu}\left(\frac{\lambda_a}{2}\right)_{\alpha\beta}\Psi^{\beta}_{j} \tag{4.5}$$

$$\mathcal{L}^{\mathrm{gg}}_{\mathrm{int}} = -\frac{g}{2}f^{abc}(\partial^{\mu}A^{\nu}_{a} - \partial^{\nu}A^{\mu}_{a})A_{b,\mu}A_{c,\nu} - \frac{g^2}{4}f^{abc}f_{ade}A^{\mu}_{b}A^{\nu}_{c}A^{d}_{\mu}A^{e}_{\nu} \tag{4.6}$$

$$\mathcal{L}^{\mathrm{FPg}}_{\mathrm{int}} = gf_{abc}(\partial_{\mu}\bar{\varphi}^{a})\varphi^{b}A^{c}_{\mu} \tag{4.7}$$

where g is the strong coupling constant, f_{abc} and λ_a are the structure constants and the Gell-Mann matrices of the SU(3) group, respectively. The interaction Lagrangian in equation 4.6 gives rise to triplet and quartic self-interactions between the gluons which lead to asymptotic freedom [16].

The β function which controls the renormalization scale dependence of the effective

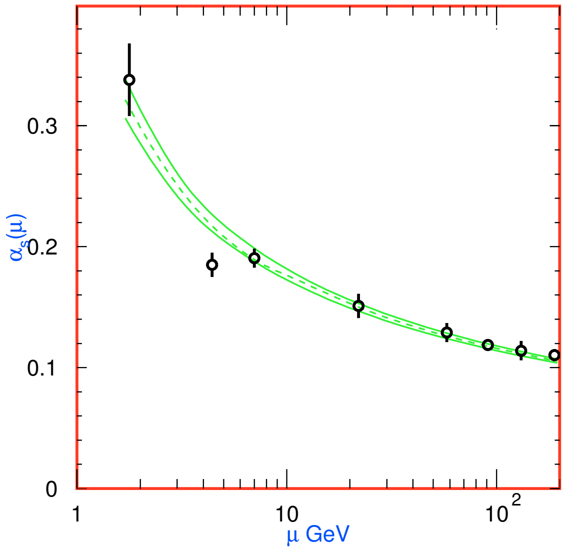


Figure 4.2.: Energy scale dependence of α_s [19].

QCD coupling $\alpha_s = g^2/4\pi$ [18] is given by

$$\mu\frac{\partial\alpha_s}{\partial\mu} = 2\beta(\alpha_s) = -\frac{\beta_0}{2\pi}\alpha_s^2 - \frac{\beta_1}{4\pi^2}\alpha_s^3 - \frac{\beta_2}{64\pi^3}\alpha_s^4 - \cdots \tag{4.8}$$

$$\beta_0 = 11 - \frac{2}{3}n_f\,,$$

$$\beta_1 = 51 - \frac{19}{3}n_f$$

where n_f is the number of quarks which are accessible at the energy scale μ. β_0, β_1,β_2 are coefficients. The solution of equation 4.8 contains an integration constant which must be determined from experiment, together with the quark masses. The currently accepted value of the strong coupling constant at $\mu = M_{Z^0}$ is $\alpha_s(M_Z) = 0.1176 \pm 0.002$ [18]. As can be seen from figure 4.2 the coupling constant $\alpha_s \to 0$ as $\mu \to \infty$, meaning when the quarks and gluons are probed at high energies they behave like free particles. Thus for collisions with large momentum transfer (Q^2) it is possible to use the perturbation theory to calculate the cross-sections approximately. Because of the large momentum transfer, such events are also called hard-scattering events.

Richard Feynman invented a method known as *Feynman Diagrams* to simplify the representation of the terms in the perturbative QED calculations as well as providing a recipe for calculating them. Table 4.1 shows the basic propagators and tree level vertices of the Feynman diagrams [15]. It is possible to combine these vertices and propagators to construct the interaction diagrams such as the so-called *Golden Decay Channel* of the Higgs boson as shown in figure 4.3. Complicated Feynman diagrams correspond to higher order terms in the perturbation theory. Monte-Carlo generators usually calculate these diagrams in the leading order. Currently only very few MC generators can calculate the diagrams at next-to-leading order.

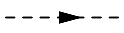
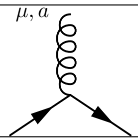
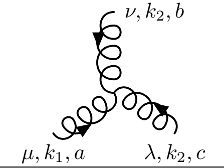
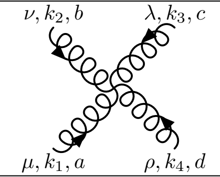
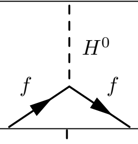
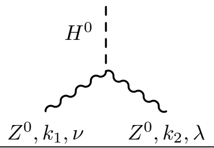
Fermion		$i\frac{\not p + m}{p^2 - m^2}$
Gluon		$i\frac{-g^{\mu\nu}\delta^{ab}}{k^2}$
Vector Boson $(W^\pm, Z^0, \gamma)$		$i\frac{-g_{\mu\nu} + k_\mu k_\nu/m_V^2}{k^2 - M_V^2}$
Higgs Boson		$\frac{i}{p^2 - m_H^2}$
Quark-gluon		$-ig_s\frac{\lambda^a}{2}\gamma\mu$
Triple-gluon		$-g_s f_{abc}\Big[g_{\mu\nu}(k_1 - k_2)_\lambda + g_{\nu\lambda}(k_2 - k_3)_\mu + g_{\lambda\mu}(k_3 - k_1)_\nu\Big]$
Quadruple-gluon		$-ig_s^2\Big[f_{abe}f_{cde}(g_{\mu\lambda}g_{\nu\rho} - g_{\mu\rho}g_{\nu\lambda}) + f_{ade}f_{bce}(g_{\mu\nu}g_{\lambda\rho} - g_{\mu\lambda}g_{\nu\rho}) + f_{ace}f_{dbe}(g_{\mu\rho}g_{\nu\lambda} - g_{\mu\nu}g_{\lambda\rho})\Big]$
Higgs-fermion		$\frac{ig}{\cos\theta_W}M_Z g_{\nu\lambda}$
Higgs-Z^0		$-i\frac{g}{2}\frac{m_f}{M_W}$

Table 4.1.: Some propagators and tree level vertices in Feynman Diagrams.

Since hadrons are composed of quarks and gluons, total hadronic cross-section can be obtained by weighting the sub-process cross-sections by the parton distribution functions (PDFs) when perturbative QCD is applicable. The parton distribution functions $f(x, Q^2)$ describe the probability of finding a parton inside the hadron carrying the total momentum fraction x when probed at the momentum scale Q. Thus in perturbative QCD, the leading order total cross-section for hadron-hadron interactions can be written as

$$\sigma = \sum_{i,j,k} \iiint \mathrm{d}x_1\, \mathrm{d}x_2\, \mathrm{d}\hat{t}\, \hat{\sigma}_{ij}^k(\hat{s}, \hat{t}, \hat{u}) \times f_{i,1}(x_1, Q^2) f_{j,2}(x_2, Q^2), \tag{4.9}$$

where $\hat{\sigma}_{ij}^k$ is the hard scattering cross-section for the k^{th} possible sub-process between partons i and j. The PDFs of the incoming hadrons are given by $f(x, Q^2)$. The Mandel-

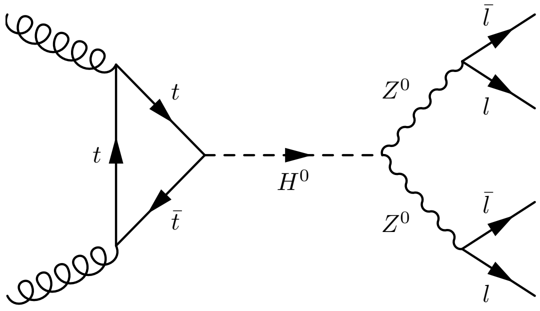


Figure 4.3.: Feynman Diagram of the so-called Golden Decay Channel of the Higgs boson. The Higgs boson (H^0) is produced from the fusion of t-quark with $\bar{t}$-quark and decays to two Z^0 bosons which in turn each decay to a lepton and an anti-lepton.

stam variables $\hat{s}, \hat{t}, \hat{u}$ satisfy $\hat{s} + \hat{t} + \hat{u} = 0$ and $\hat{s} = x_1, x_2 s$ for massless particles. Using equation 4.9 it is possible to write the hard-scattering cross-section above some $p_{\mathrm{T}_{\mathrm{min}}}$ as [20]

$$\sigma_{\mathrm{hard}}(p_{\mathrm{T}_{\mathrm{min}}}) = \int_{{p_{\mathrm{T}_{\mathrm{min}}}}^2}^{s/4} \frac{\mathrm{d}\sigma}{\mathrm{d}{p_{\mathrm{T}}}^2} \mathrm{d}{p_{\mathrm{T}}}^2 \tag{4.10}$$

which diverges like $\alpha_s^2({p_{\mathrm{T}}}^2)\mathrm{d}{p_{\mathrm{T}}}^2 / {p_{\mathrm{T}}}^4$ as the events get softer i.e. $p_{\mathrm{T}_{\mathrm{min}}} \to 0$. Since perturbative QCD calculations are not applicable for low momentum transfer processes, other models based on phenomenological approaches, with different accuracies such as the *Eikonal*, *Regge poles* or *Donnachie-Landshoff model for the "soft pomeron"* are used [21]. However none of these models can describe the soft interactions completely and very few MC generators can generate soft scattering events. Some examples of such generators are PYTHIA, using an approach combining Regge theory [22] and "eikonalized" QCD models [23], PHOJET, using the "Two Component Dual Parton Model [24]" and EPOS using a phenomenological approach based on the parton model [25].

4.2. Event Signatures

The particle interactions are usually classified with respect to their signatures in the detector. In the elastic events the momentum transfer is too small to break-up incoming particles, A and B, and slightly alters their trajectories which create hits at very high-η values if at all and represented by

$$\mathrm{A} + \mathrm{B} \to \mathrm{A} + \mathrm{B}.$$

In the single-diffractive events, one of the incoming particles is dissociated through exchange of a colorless and flavorless particle and creating particle showers, X, at one side of the η–ϕ plane. It is represented by

$$\begin{aligned} \mathrm{A} + \mathrm{B} &\rightarrow \mathrm{X}_1 + \mathrm{B} \\ \mathrm{A} + \mathrm{B} &\rightarrow \mathrm{A} + \mathrm{X}_2 \end{aligned}$$

In double-diffractive events, both particles are dissociated, creating activity in both sides of the η–ϕ plane with some region with no activity in between and represented by

$$\mathrm{A} + \mathrm{B} \rightarrow \mathrm{X}_1 + \mathrm{X}_2.$$

It is also possible to observe events with higher diffractive dissociations where there are more showers in the η–ϕ plane with rapidity gaps in between them.

In non-diffractive events both incoming particles are dissociated through exchange of a colored particle creating activity in the whole η–ϕ plane. Figure 4.4 shows the illustrations of events types in η–ϕ plane. Minimum bias events are, as the name suggests, the events that are recorded with as little bias from the trigger as possible. They are usually associated with non-diffractive events, however sometimes double-diffractive events are also included in the definition since it is experimentally rather hard to completely separate the double-diffractive events and non-diffractive events.

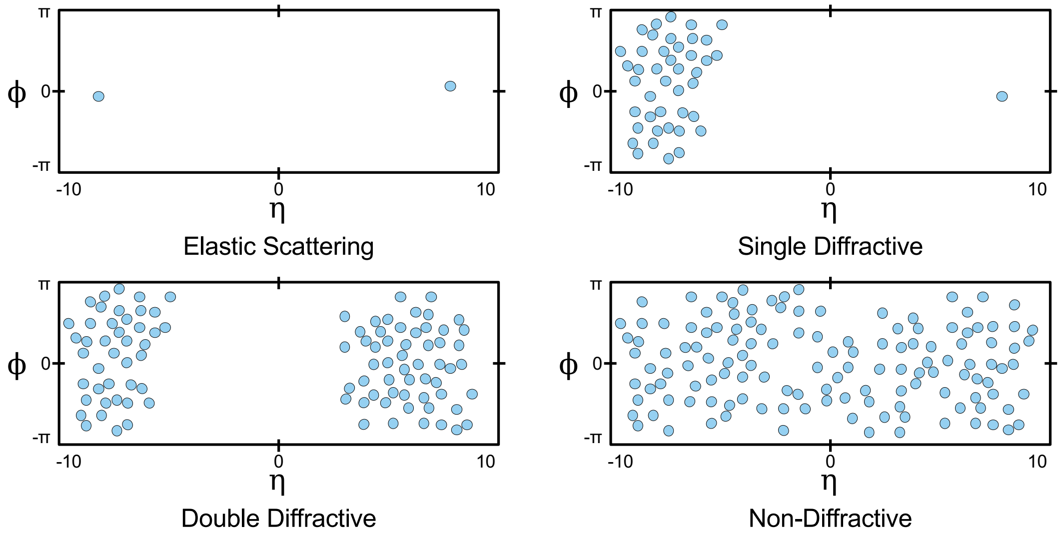


Figure 4.4.: Event topologies in η–ϕ space.

Following the event types, the total cross-section is usually studied by dividing it into elastic, single-diffractive, double-diffractive and non-diffractive parts as

$$\sigma_{\mathrm{tot}} = \sigma_{\mathrm{el}} + \sigma_{\mathrm{sd}} + \sigma_{\mathrm{dd}} + \sigma_{\mathrm{nd}}.$$

4.3. Reggeons and Pomerons

A Reggeon is a pole in the partial wave in the t-channel of the scattering process [26]. The partial wave scattering amplitude of spinless particles exchanged in the t-channel can be written as [27]

$$A(s,t) = 16\pi \sum_{l=0}^{\infty} (2l+1) A_l(t) P_l(z_t) \tag{4.11}$$

where $P_l(z_t)$ is the Legendre polynomial of the first kind, $A_l(t)$ is the partial wave amplitude and

$$z_t = \cos\theta_t = 1 + \frac{2s}{t - 4m^2}, \tag{4.12}$$

in the physical region of the t-channel where $t > 4m^2$, $s < 0$ and $|z_t| \leq 1$. As $s \to \infty$, z_t becomes proportional to s and since for large z, $P_l(z) \sim z^l$, the series in equation 4.11 diverges. However, this sum can be converted in to an integral through the Sommerfeld-Watson transform

$$A^{\pm}(z_t, t) = 8\pi i \int_C dl (2l+1) A^{\pm}(l,t) \frac{P_l(-z_t) \pm P_l(z_t)}{\sin(\pi l)} \tag{4.13}$$

where C is the contour surrounding the real axis from 0 to ∞, $A(l,t)$ is the partial wave amplitude coinciding with the $A_l(t)$ at the physical values of complex angular momentum l and $\pm$ denotes the even-ness of l. If one of the partial wave amplitudes has a pole in the complex l plane such that

$$A(l,t) \approx \frac{G(t)}{l - \alpha_{\mathbb{R}}(t)} \tag{4.14}$$

then the function $\alpha_{\mathbb{R}}(t)$ is said to define the Reggeon trajectory [26]. It has the form

$$\alpha_{\mathbb{R}}(t) = \alpha_{\mathbb{R}}(0) + \alpha'_{\mathbb{R}}(0)t, \tag{4.15}$$

where $\alpha_{\mathbb{R}}(0)$ is the intercept of the Reggeon and the $\alpha'_{\mathbb{R}}(0)$ is its slope. Then the Pomeron is defined as *a Reggeon with the intercept close to 1* or

$$\alpha_{\mathbb{P}}(0) - 1 = \Delta \leq 1. \tag{4.16}$$

The Pomeron exchange explains the diffractive events at high energies surprisingly well [28]. Because of this, they are usually used in the representation of diffractive events similar to Feynman diagrams as in figure 4.5. The detailed explanation of the Pomerons and the Regge theory is available elsewhere [26, 27, 29].

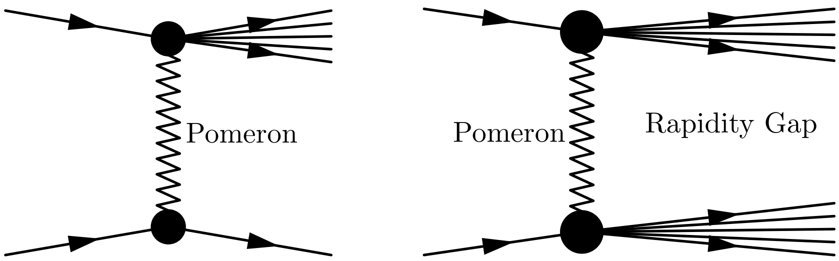


Figure 4.5.: A typical representation of pomeron exchange in single-diffractive and double-diffractive events. Left figure represents single-diffractive events and right figure represents double-diffractive events.

4.4. Monte-Carlo Event generators

Monte Carlo (MC) technique is a statistical method using a probabilistic approach. Most likely the first example of the method would be the *Buffon's needle* experiment. In this experiment a needle is thrown on a plane with parallel equidistant lines. The probability of the needle falling on a line is proportional to the distance between the lines, length of the needle and π. Thus by throwing the needle many times and counting the number of times that it crossed a line, one can estimate the value of π. Although this method is not the best method to calculate the value of π, it is possible to estimate it within a desired degree of accuracy with an appropriate number of throws. Since the accuracy is proportional to the number of throws, computers using random numbers to simulate the throwing would be an ideal choice for such a task. S. Ulam, J. von Neumann and N. Metropolis were, in 1949, the first to implement the MC technique on a computer to estimate the distance that a neutron from a radioactive decay would travel in matter [30]. Since then the power of the method has been understood and widely used in almost all branches of science.

The MC method is usually applied to problems where the calculation of the exact answer is unfeasible or impossible. Integral evaluations are one of the basic applications of the MC method since one can write

$$\lim_{N\to\infty} \frac{|a-b|}{N} \sum_{i=0}^{N} f(\mathbb{R}_i(a,b)) = \int_a^b f(x)\mathrm{d}x \tag{4.17}$$

where $f(x)$ is the integrand, $\mathbb{R}_i(a,b)$ is a random point in the integration space and N is the number of samples. The difference between the integral and the sum decreases as the number of samples increases, independent of the dimensions of the integration space. However CPU time needed for the calculation of integral through analytic integration methods, if possible at all, grows exponentially by the dimension of the integration space. Considering the probabilistic nature of quantum mechanics and multi-dimensional integrals it contains, the MC method is a natural match for the problems in high energy physics and found various applications in fields such as calculation of Feynman diagrams, event generation and detector simulation.

The event generators try to simulate events in accelerators in a virtual world, as well as our understanding of the processes permits. However we still do not know how to describe nor how to calculate all the stages of an event formation, thus event generators use computations of the parts which can be calculated from first principles and use models for the rest. Typical steps in the generation of a *pp* or $p\bar{p}$ event can be outlined as follows [31]:

- Select a parton from both incoming particles and generate the selected interaction.
- Process the decays of the short lived particles such as Z^0, $W^{\pm}$, t-quark coming out of the selected interaction.
- Since colored partons take part in the collision they can go through bremsstrahlung. Bremsstrahlung of the incoming partons are called Initial-State Radiation (ISR) and outgoing partons are called Final-State Radiation (FSR). Simulate these effects.
- Repeat the same steps for other partons which could also undergo collisions.
- Only a fraction of the incoming energy is used in collisions, remaining partons (beam remnants) carry the rest of the energy and color compensating for the colors of interacting partons. All these partons and those from the hard collision move apart. Due to confinement they can only be observed as color singlets thus models for hadron formation must be employed.
- Short-lived particles such as B mesons and τ leptons that cannot be seen in detectors should be allowed to decay.

Similar steps can be applied to e^-e^+ or e^-p collisions.

The first step involves calculation of matrix elements which correspond to Feynman diagrams for many hard processes. Unfortunately, often these elements are available only at the lowest order. Thus several MC programs specialize in calculation of a selected set of matrix elements and pass their calculations to the generic event generators for the remaining steps. An alternative approach is called Parton Showers and works by generating complex $2 \to n$ events by factorizing them in terms of simple expressions such as $q \to qg$, $g \to gg$ and $g \to q\bar{q}$. Although it is not exact, it can provide sensible approximations.

Incoming and outgoing partons are allowed to emit bremsstrahlung. The emissions get harder as partons approach the main interaction and get softer as they recede. The emissions in ISR and FSR are bounded by a lower limit and upper limit and they are usually modeled by the parton shower approach. Figure 4.6 shows a representation of the ISR.

As time progresses, the partons created in the interaction, ISR and FSR move away from each other. But they are related with each other by color connections. Since such processes can not be described by perturbative QCD, various models are used.

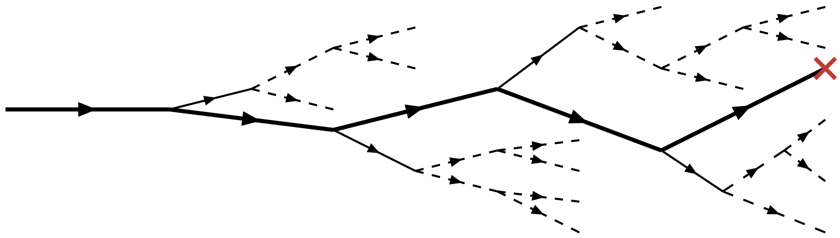


Figure 4.6.: A representation of ISR. The thick line is the main parton taking part in the main interaction (the X mark). Thin lines are partons that can not be recombined and dashed lines are further fluctuations that may or may not recombine. FSR is similar [31].

One of these models is *String Fragmentation*. In string fragmentation the partons are assumed to be connected to each other through a flux tube. As partons move away the tube is stretched until it breaks down in the middle, producing either a $q\bar{q}$ pair or diquark-antidiquark pair. The process continues until all strings are stable thus forming hadrons. The process is called fragmentation/hadronization and the model requires several parameters to be defined from the data. After the hadronization, short-lived hadrons are decayed depending on their decay channels and widths.

After an event generator finishes processing an event, the results can be passed through a detector simulation MC to imitate the measurements of a detector. These measurements (hits) can be analyzed through reconstruction and analysis programs to study various aspects of the experiment such as the trigger and data acquisition systems, efficiencies, acceptances, physics signals and possible backgrounds.

4.5. PYTHIA

PYTHIA is one of the best known and widely used event generators [32]. It has many hard QCD processes implemented and supports Super-symmetry (SUSY) and several Beyond Standard Model (BSM) physics. It uses the Lund string model for shower evolution and can read events dumped in the Les Houches accord [33] format. It is written in FORTRAN (up to PYTHIA 6.x). The recently released versions 8.x is written in C++. PYTHIA has many parameters to alter the operation of the program, giving advanced users a lot of flexibility. The default settings, on the other hand, enable novice users to run the program just with a couple of lines of code.

In PYTHIA the total cross-section of hadron-hadron interactions are calculated using the formula ([34])

$$\sigma_{\text{tot}}^{\text{AB}}(s) = X^{\text{AB}} s^{\epsilon} + Y^{\text{AB}} s^{-\eta} \tag{4.18}$$

where $s = E_{\text{cm}}^2$ and X^{AB} and Y^{AB} correspond to pomeron and reggeon terms depending on the initial states of incoming particles A and B. The powers are taken as $\epsilon = 0.0808$ and $\eta = 0.4525$.

The elastic cross-section is calculated using the formula

$$\sigma_{\mathrm{el}} = \frac{\sigma_{\mathrm{tot}}^2}{16\pi B_{\mathrm{el}}} \tag{4.19}$$

where the elastic slope parameter B_{el} is parameterized as

$$B_{\mathrm{el}} = B_{\mathrm{el}}^{\mathrm{AB}}(s) = 2b_{\mathrm{A}} + 2b_{\mathrm{B}} + 4s^{\epsilon} - 4.2 \tag{4.20}$$

with the constants $b_{\mathrm{A,B}}$ being $b_{\mathrm{p}} = 2.3$, $b_{\pi,\rho,\omega,\phi} = 1.4$, $b_{J/\psi} = 0.23$ for $p, \pi, \rho, \omega, \phi$ respectively.

The diffractive cross-sections are given by

$$\frac{\mathrm{d}\sigma_{\mathrm{sd(XB)}}\,(s)}{\mathrm{d}t\,\mathrm{d}M^2} = \frac{g_{3\mathbb{P}}}{16\pi}\beta_{\mathrm{A}\mathbb{P}}\beta_{\mathrm{B}\mathbb{P}}^2\frac{1}{M^2}\exp(tB_{\mathrm{sd(XB)}})F_{\mathrm{sd}} \tag{4.21}$$

$$\frac{\mathrm{d}\sigma_{\mathrm{sd(AX)}}\,(s)}{\mathrm{d}t\,\mathrm{d}M^2} = \frac{g_{3\mathbb{P}}}{16\pi}\beta_{\mathrm{A}\mathbb{P}}^2\beta_{\mathrm{B}\mathbb{P}}\frac{1}{M^2}\exp(tB_{\mathrm{sd(AX)}})F_{\mathrm{sd}} \tag{4.22}$$

$$\frac{\mathrm{d}\sigma_{\mathrm{dd}}\,(s)}{\mathrm{d}t\,\mathrm{d}M_1^2\,\mathrm{d}M_2^2} = \frac{g_{3\mathbb{P}}^2}{16\pi}\beta_{\mathrm{A}\mathbb{P}}\beta_{\mathrm{B}\mathbb{P}}\frac{1}{M_1^2}\frac{2}{M_2^2}\exp(tB_{\mathrm{dd}})F_{\mathrm{dd}}. \tag{4.23}$$

Equations 4.21 and 4.22 represent the single-diffractive events where the first or second proton is dissociated, respectively. The double-diffractive cross-section is calculated with equation 4.23. M represents the mass of the diffractive system and coupling terms $\beta_{\mathrm{A}\mathbb{P}}$ are related to the pomeron term in equation 4.18. They are selected such that $\beta_{\mathrm{A}\mathbb{P}}\beta_{\mathrm{B}\mathbb{P}} = X^{\mathrm{AB}}s_{\mathrm{ref}}^{\epsilon}$ where $s_{\mathrm{ref}}^{\epsilon} = \sqrt{20}$ GeV. The triple pomeron coupling is taken as $g_{3\mathbb{P}} \approx 0.318\,\mathrm{mb}^{1/2}$. F_{sd} and F_{dd} are fudge factors introduced in order to obtain a sensible behavior in the whole phase space. These equations are integrated over the full phase space at different center-of-mass energies and the results are parameterized. The non-diffractive cross-section σ_{nd} is found by subtracting diffractive and elastic cross-sections from the total cross-section given by equation 4.18. Details about calculations of diffractive cross-sections and the parameterization are available elsewhere [32, 35]. Starting from the PYTHIA 8 version 8.130, a new mechanism for diffractive event production, making use of the pomeron PDFs, is introduced. For the details of the mechanism see [36].

For some values of $p_{\mathrm{T_{min}}}$ and s, σ_{hard} given in equation 4.10 might exceed σ_{tot}. Although this might seem unphysical it can be interpreted as the total cross-section of multiple parton interactions rather than one interaction [20]. In PYTHIA the average number of parton-parton interactions in an event is found by using the relation $\langle n\rangle = \sigma_{\mathrm{hard}}/\sigma_{\mathrm{nd}}(s)$, where σ_{nd} is the non-diffractive cross-section. Multiple parton interactions are modeled with or without a dependence on the transverse distance between the incoming partons, the impact parameter. The basic model assumes no impact parameter dependence. An advanced model uses different matter distribution models, one

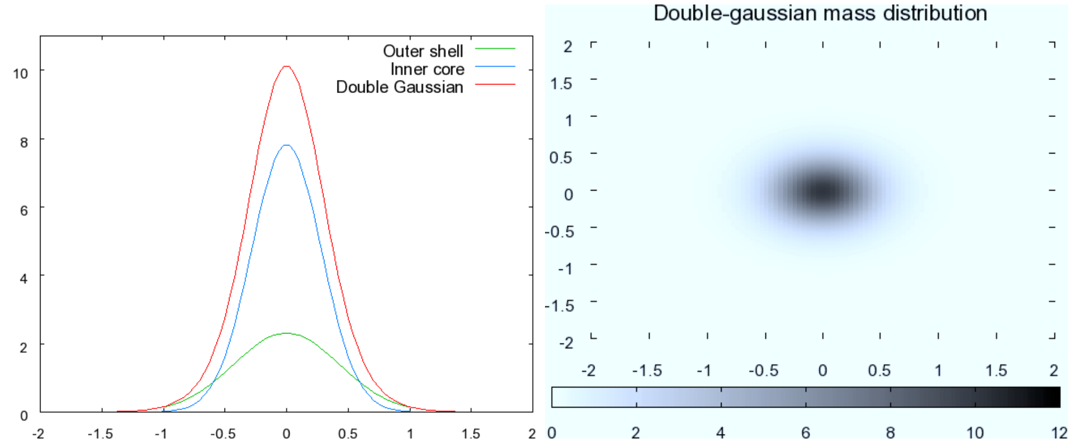


Figure 4.7.: An example of double-Gaussian matter distribution given in equation 4.24 with parameter settings $a_1 = 0.6$, $a_2 = 0.4$ and $\beta = 0.5$. Left plot shows Gaussian distributions representing combined, inner-core and outer-shell. Right plot shows the distribution viewed from above.

of which is a co-centric double-Gaussian with individual mass densities given by

$$\rho(r) \propto \frac{1-\beta}{a_1^3} \exp\left\{-\frac{r^2}{a_1^2}\right\} + \frac{\beta}{a_2^3} \exp\left\{-\frac{r^2}{a_2^2}\right\} \tag{4.24}$$

In this equation a_1 and a_2 correspond to the radii of the outer and inner Gaussians, respectively, and β is the fraction of the total hadronic matter inside the Gaussian. Figure 4.7 shows an example distribution with default parameters. Another impact parameter dependent model uses a matter overlap profile in the form of

$$\mathcal{O}(b) \propto \exp(-b^d) \tag{4.25}$$

where b is the impact parameter of the collision and d is the model parameter. Details about the models and event generation are available in reference [37].

The divergence of the hard cross-section at low-p_{T} is regulated either by a sharp or a gradual cutoff [20]. Gradual regularization is done by multiplying the matrix elements with the factor

$$\frac{\alpha_s({p_{\mathrm{T}_0}}^2 + {p_{\mathrm{T}}}^2)}{\alpha_s({p_{\mathrm{T}}}^2)} \frac{{p_{\mathrm{T}}}^4}{({p_{\mathrm{T}_0}}^2 + {p_{\mathrm{T}}}^2)^2} \tag{4.26}$$

With this approach the perturbative QCD cross-section is recovered when $p_{\mathrm{T}} \gg p_{\mathrm{T}_0}$ and strongly damped otherwise. Hard interaction cut-off, $p_{\mathrm{T}_{\mathrm{min}}}$, and the regulation parameter, p_{T_0}, are chosen as energy dependent values, given by the formulae

$$p_{\mathrm{T}_{\min}}(E_{\mathrm{cm}}) = p_{\mathrm{T}_{\min}}^{ref} \left(\frac{E_{\mathrm{cm}}}{E_{\min}^{ref}} \right)^{E^{pow}} \tag{4.27}$$

$$p_{\mathrm{T}_0}(E_{\mathrm{cm}}) = p_{\mathrm{T}_0}^{ref} \left(\frac{E_{\mathrm{cm}}}{E_0^{ref}} \right)^{E^{pow}} \tag{4.28}$$

In these equations, E_0^{ref} and $E_{\min}^{ref}$ represent the scales where p_{T_0} and $p_{\mathrm{T}_{\min}}$ are equal to their respective reference scales $p_{\mathrm{T}_0}^{ref}$ and $p_{\mathrm{T}_{\min}}^{ref}$. E^{pow} is common in both equations and defines the extrapolation slope. These parameters should be determined by data. $p_{\mathrm{T}_{\min}}^{ref}$, $p_{\mathrm{T}_0}^{ref}$ and E^{pow}, together with the matter distribution parameters, are important parameters in the tuning process.

In PYTHIA initial state showers do not evolve below $Q_0 = 1$ GeV. However this is compensated by assigning a primordial k_{T} to shower initiators as a source of softer activity. There are different choices for the distributions of primordial k_{T}. As an example for PYTHIA 6 it is

$$e^{(-k_{\mathrm{T}}^2/\sigma^2)} k_{\mathrm{T}} \mathrm{d}k_{\mathrm{T}} \tag{4.29}$$

with an upper cut-off. On the other hand for PYTHIA 8 it is a Normal Distribution with a width

$$\sigma = \frac{(\sigma_{\mathrm{soft}} Q_{\mathrm{half}} + \sigma_{\mathrm{hard}} Q)}{Q_{\mathrm{half}} + Q} \frac{m}{m_{\mathrm{half}}} \tag{4.30}$$

where Q is the scale of the hardest process, m is the mass of the system and σ_{hard}, σ_{soft}, Q_{half} and m_{half} are parameters.

Explaining the capabilities and models in PYTHIA is beyond the scope of this thesis. For detailed discussions and explanations of the models used in PYTHIA 6 and PYTHIA 8, see [32, 36, 38] and the references therein.

4.6. PHOJET

The Dual Parton Model (DPM) is a phenomenological model which can calculate both elastic processes and inelastic processes [39]. PHOJET uses the Two-Component DPM, which connects both soft and hard processes by an unitarization scheme with an extension to photons [40]. It does not have user tunable parameters since all free parameters of the model are defined from the data. Aside from elastic, single, double and non-diffractive events, it can also generate central-diffractive events where there is an activity in the central region separated by a rapidity gap from the activity on both sides of the $\eta - \phi$ plane.

Cross-sections for the hard processes where the transverse momentum cut-off is much greater than the QCD scale, i.e. $p_{\mathrm{T}}^{\mathrm{cutoff}} \gg \Lambda_{\mathrm{QCD}}$, are calculated using

$$\sigma_{\mathrm{AB}}^{\mathrm{hard}}(s, p_{\mathrm{T}}^{\mathrm{cutoff}}) = \int \mathrm{d}x_1 \, \mathrm{d}x_2 \, \mathrm{d}\hat{t} \sum_{i,j,k,l} \left(\frac{1}{1+\delta_{k,l}} f_{i|A}(x_1, \mu^2) f_{j|B}(x_2, \mu^2) \right. \\ \left. \times \frac{\mathrm{d}\sigma_{\mathrm{i,j \to k,l}}^{\mathrm{QCD}}(\hat{s}, \hat{t})}{\mathrm{d}\hat{t}} \Theta(p_{\mathrm{T}} - p_{\mathrm{T}}^{\mathrm{cutoff}}) \right) \tag{4.31}$$

where i, j, k and l are possible parton configurations and $\hat{s}, \hat{t}$ and $\hat{u}$ are Mandelstam variables of the partonic scattering process. Born graph cross-section for hadron-hadron scattering is approximated by the formula [41]

$$\sigma_{\mathrm{AB}}^{\mathrm{Born}} = \sigma_{\mathrm{AB}}^{\mathbb{R}} + \sigma_{\mathrm{AB}}^{\mathbb{P}_s} + \sigma_{\mathrm{AB}}^{\mathrm{hard}} \tag{4.32}$$

with a super-critical soft pomeron term

$$\sigma_{\mathrm{AB}}(s)^{\mathbb{P}_s} = g_{\mathrm{A}\mathbb{P}}^0 g_{\mathrm{B}\mathbb{P}}^0 \left(\frac{s}{s_0} \right)^{\Delta_{\mathbb{P}_s}} \quad , \quad \Delta_{\mathbb{P}_s} = \alpha_{\mathbb{P}_s}(0) - 1$$

and reggeon term

$$\sigma_{\mathrm{AB}}(s)^{\mathbb{R}} = g_{\mathrm{A}\mathbb{R}}^0 g_{\mathrm{B}\mathbb{R}}^0 \left(\frac{s}{s_0} \right)^{\Delta_{\mathbb{R}}} \quad , \quad \Delta_{\mathbb{R}} = \alpha_{\mathbb{R}}(0) - 1$$

where $\alpha(0)$ are the Regge trajectories with largest intercept and g^0 terms are the pomeron and reggeon coupling constants. The coupling constants, $g_{\mathbb{R}}$ and $g_{\mathbb{P}}$, and the intercept terms, $\alpha_{\mathbb{R}}$ and $\alpha_{\mathbb{P}}$, are determined from fits to $p\bar{p}$ cross-section measurements. PHOJET generates the main interaction and uses PYTHIA for fragmentation and hadronization. The use of PHOJET and its model is described in references [40, 41] and the references therein.

4.7. EPOS

EPOS is a hadronic interaction model with a multiple scattering approach based on partons and pomerons [25]. The EPOS generator is used for heavy-ion collisions and cosmic-ray interactions and can describe the existing minimum bias data quite accurately [42–45].

EPOS stands for **E**nergy conserving quantum mechanical multiple scattering approach, based on **P**artons (parton ladders) **O**ff-shell remnants and **S**plitting of parton ladders. The parton ladder structure is composed of two parts, a hard part including the space-like initial state cascade and the hard interaction, and a soft part which is a purely phenomenological object parameterized in Regge pole fashion [25]. In addition to the parton ladder, remnants of the incoming particles are included in calculations as colorless and usually off-shell objects. The energy-momentum sharing is taken into consideration at the parton-ladder level which is a unique feature of EPOS. Another unique feature is the use of the same formulae for the partial-cross-section calculation and parton generation.

EPOS also has several drawbacks. Probably the most notable one is the lack of a manual. Another one is that it does not have process selection and certain processes are not implemented yet. However these problems are expected to be solved in the near future and it can be used for the minimum bias predictions at the LHC already. The details of the EPOS model are available in [25, 43, 46] and the references therein.

5. Monte-Carlo Tuning with Genetic Algorithms

Due to the uncertainties in the underlying models, Monte Carlo event generators usually have parameters which can be tuned to describe the data better. The tuning process involves comparison of MC generated distributions with data and adjustment of the parameters until they match.

There are two important issues in MC tuning. One of them is the generation of distributions. The generation of MC distributions should be done as close to the experimental conditions as possible. The event generation and selection to describe the experimental conditions is the most important and most crucial step of any tuning process, since it defines the optimum parameter set. That is, the optimum parameter set for a crude implementation of event selection and experimental conditions can be a very different than from a detailed one.

Another important thing in tuning is the search method used in finding the optimum set. The parameters are altered within given ranges which form a volume with the number of dimensions equal to the number of parameters to tune. Each parameter set corresponds to a point within this volume. Usually a MC generator has several parameters to tune which makes the volume of the parameter space

$$V = \prod_{i=0}^{n} r_i$$

where r is the range of parameter and n is the number of parameters to tune. Thus as the number of parameters and their ranges increase it becomes harder to find an optimum parameter set. This part is essentially a search problem in a multi dimensional space.

There are various methods used in search problems. In general, the following list of methods are used for steering the search of a good parameter set.

- manually by eye
- fitting distributions to the data
- brute force
- parameterization
- Genetic Algorithms.

Probably the most basic approach to tuning a generator would be changing some parameters and looking at the distributions. In this method, the user steers the parameters depending on his or her observations of the distributions. Although this method produced some of the most commonly used tunes to date, it requires expertise and depends on the judgement of the person. Moreover it becomes a cumbersome process as the number of parameters or distributions increases.

Another approach is trying to fit the generated distributions to the data using a minimization package such as Minuit [47]. This method can work, but it has very little potential to be run in parallel and for each point in the parameter space, evaluation of several points is required for calculating the derivatives. Considering that the evaluation of a point may take a long time, it quickly becomes unfeasible for large number of parameters.

The brute force approach is based on random evaluation of the points in the parameter space with the hope that one of them will be a good match to data. It can be parallelized and might work for one or two parameters with small ranges. However it quickly becomes unfeasible for any realistic number of parameters.

The idea of the parameterization of the generator response is based on the evaluation of randomly selected points in the parameter space and then bin-by-bin parameterizing the resulting distributions using a polynomial function of the parameters. A tool called Professor [48–50] using the approach based on the ideas in [51] works in this manner. In Professor, generator response for each distribution is bin-by-bin parameterized using a second or third order polynomial function of tuned parameters by singular value decomposition method. Then a χ^2 minimization is done using these parameterized distributions to find the optimum values of the parameters. A detailed description of Professor can be found in [52] and the references therein.

The parameterization approach provides a way of doing systematic tuning and reducing the subjectivity. However it assumes the generator response can be sufficiently approximated by a polynomial function of the input parameters. The order of this polynomial cannot be very high since the minimum number of evaluations required for a singular value decomposition is

$$\sum_{n=0}^{n=N} \frac{(n+P-1)!}{n!(P-1)!}$$

where N is the order of the polynomial and P is the number of parameters.

Parameterizing the generator response permits the use of analytic minimization techniques, however if the generator response is more complicated than the polynomial, the minima of the function set might not necessarily be even a local minimum.

The approach applied to the tuning problem in this thesis is to use Genetic Algorithms to steer the optimum parameter search. Genetic Algorithms provide a compromise between the analytic search and random sampling. They do not make any assumptions about the functional dependence of the generator response on the parameters and are less likely to converge to a local minimum. In this respect it is a combination of all previous methods.

The data sets, event generation and selection conditions used in this thesis are described in section 5.1. A basic description of the Genetic Algorithms and their application to MC tuning are discussed in sections 5.2 and 5.3, respectively. Sections 5.4 and 5.5 describe the tuning of PYTHIA for minimum bias events and the results of tuning.

5.1. Data

There are several measurements made by different groups at various energies using minimum bias events. Typical minimum bias distributions are p_T, N_ch, $dN_\mathrm{ch}/d\eta$, and $< p_\mathrm{T} >$. These measurements are used for tuning the parameters of MC generators. Most of the time the experimental groups define the minimum bias events depending on the properties of their detectors. Such definitions slightly vary and create differences between data sets. Also some data are corrected for inefficiencies, acceptances, resolution and other effects while the others are not. Furthermore for some data it is not clear in the article how the measurement is done. Such problems usually raise reliability or reproducibility problems.

Some of the minimum bias measurements are rather old and the data are published as plots instead of numerical values. There is an on-line database [53] where the experimental HEP data are collected whenever it is possible. Unfortunately this database does not contain all the data necessary for minimum bias measurements. The data from this database has been used whenever it was possible. For data which are available only as plots, a special tool was prepared to make measurements on the plots with high accuracy.

The definition of minimum bias events in the generators usually depend on the author(s) of the program. Thus to have a somewhat consistent approach between the data and the generators, aside from acceptances, a basic event selection criteria, as close as possible to the data conditions, must be implemented. Most of the time that criteria is an approximation of the trigger used by the experiment. Because of the complex relations between the parameters used in minimum bias tuning and the close ties between minimum bias and underlying event, which represents the remainder of interacting particles aside from main interaction, some underlying event and Drell-Yan data are also used in tuning. Although it is possible to find a good description of the minimum bias data without these data sets, non-minimum bias distributions are significantly distorted with respect to data.

5.1.1. CDF

The CDF Experiment is one of the two experiments at the Tevatron at Fermilab. The CDF collaboration has made several minimum bias measurements using $p\bar{p}$ data recorded at 630 GeV, 1800 GeV and 1960 GeV center-of-mass energy. The CDF detector is an azimuthally symmetric general purpose detector. Only the relevant parts of the detector that are used in analysis is briefly described here. The details can be found in ref. [54]. Figure 5.1 shows a cross-sectional view of one quadrant of the CDF Central Detector.

Two sets of scintillation counters called Beam-Beam Counters (BBC) are located at ±5.82 m away from the nominal interaction point of the CDF. They cover the pseudorapidity range $3.2 \leq |\eta| \leq 5.9$ and form the basis of the CDF "minimum bias trigger".

In the central region the Vertex Time-Projection Chambers (VTPC) lies closest to the beam pipe and is composed of 8 time-projection chambers. It covers the η ranges $3.5 \leq |\eta|$ with the inner layer and $2.6 \leq |\eta|$ with the outer layer. The VTPC provides high precision tracking information of the charged particles. It is used in measurements of the event topology and the determination of the event vertex position.

The VTPC is surrounded by the Central Tracking Chamber (CTC). The CTC is a cylindrical drift chamber immersed in a 1.5 T magnetic field and has 86 layers of wires distributed over its radial thickness. It provides precise momentum measurements in the range $-1 \leq \eta \leq 1$.

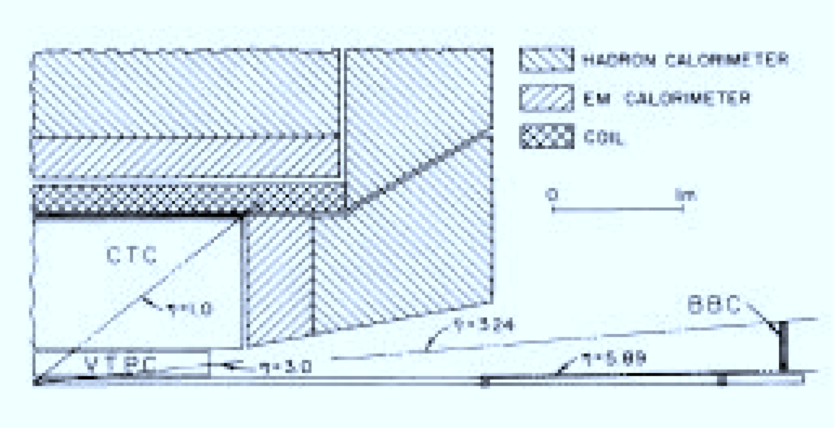


Figure 5.1.: Cross-sectional view of one quadrant of the CDF Central Detector.

In 2001, the CDF detector has been upgraded to the CDF-II, replacing all tracking detectors. The VTPC is replaced by a Silicon Vertex Detector (SVX-II). The SVX-II contains five layers of double-sided silicon micro-strip detectors between the radii of 2.4 cm to 10.7 cm. The Intermediate Silicon Layers (ISL) are double-sided silicon strips, placed at radii of 22 cm for $|\eta| \leq 1$ and at 20 cm and 28 cm for $1 \leq |\eta| \leq 2$. The Central Outer Tracker (COT), covering the radial range from 44 cm to 132 cm, replaced the CTC. The COT is a drift chamber with roughly five times the readout channels the CTC had. Figure 5.2 shows the layout of the upgraded tracker systems. The detailed description of the CDF-II detector can be found in [55, 56] and the references therein.

Transverse Momentum Distributions of Charged Particles

The CDF collaboration has done several measurements of the transverse momentum distributions of charged particles. The first measurement was done using the data collected in the Run-0 period at 630 GeV and 1800 GeV. The data used in this analysis were collected with a coincidence trigger, requiring at least one hit in the BBCs on both sides

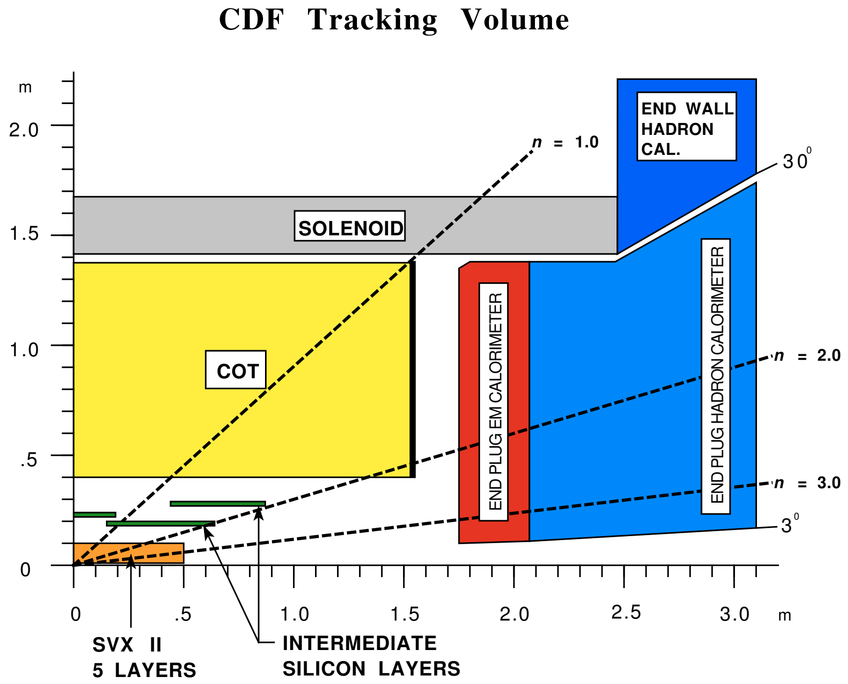


Figure 5.2.: Cross-sectional view of one quadrant of the CDF-II tracker systems. The innermost layer is the Silicon Vertex Detector (SVX-II). The Intermediate Silicon Layers (ISL) lie between the COT and SVX-II. The Central Outer Tracker (COT) forms the last part of the CDF-II tracking region.

of the detector in coincidence with the beam crossing. In order to increase the tracking efficiency and acceptance, only the events with a reconstructed vertex within 65 cm of the nominal beam crossing point was used. The events are further selected by requiring at least four charged particles in the range $|\eta| < 3$ with at least one charged particle in both the $+\eta$ and $-\eta$ hemispheres. Finally using the charged particles with $|\eta| < 1$ and $p_{\mathrm{T}} > 0.4$ GeV/c and assuming all particles to be pions, $\mathrm{E}\,\mathrm{d}^3\sigma\,/\mathrm{d}^3p$ is calculated and the normalization of the inclusive cross-section was estimated from the measurements of the UA4 experiment. The measurement was corrected for the decays of charged pions and kaons, secondary interactions, photon conversions and the decays of neutral strange particles. Further details of the analysis are available elsewhere [57].

The second measurement was done using data collected in the Run-1 period using a coincidence trigger. Additionally the selected events were required to have only one primary vertex and cosmic ray particle tracks were vetoed. In the selected events tracks which did not traverse a certain number of CTC layers were omitted from the analysis. Tracks which did not pass within 0.5 cm of the beam axis and 5 cm of the primary vertex in beam direction were labeled as secondary tracks and neglected. Finally the remaining tracks with $p_{\mathrm{T}} \geq 0.4$ GeV/c and within $|\eta| \leq 1.0$ was used in the measurement of $\mathrm{E}\,\mathrm{d}^3N\,/\mathrm{d}p^3$. Reference [58] describes the analysis details.

The third measurement was done with the CDF-II detector using data collected during the Run-2 period. The data are collected using a coincidence trigger requiring at least one charged particle hit in $3.7 < |\eta| < 4.7$ at both sides of detector in coincidence with the beam crossing. The analysis required successful primary vertex reconstruction. From the events satisfying this criteria, the tracks which had a certain minimum number of hits, originated from certain fiducial volume, with $p_{\mathrm{T}} \geq 0.4$ GeV/c and within $|\eta| \leq 1$ were used in $\mathrm{E}\,\mathrm{d}^3\sigma\,/\mathrm{d}p^3$ measurement. The details of the analysis are presented in [59].

Pseudorapidity distributions of charged particles

The Run-0 minimum-bias data at 1800 GeV and 630 GeV were also used in the measurement of the pseudorapidity distributions of charged particles. The data were collected using the BBC coincidence trigger. The events are offline selected by requiring either one the two criteria satisfied. The first critera required a minimum of four tracks in the VTPC with at least one in both η hemispheres. The second criteria required the interaction point determined with at least two tracks lying within 16 cm of the interaction point determined by the BBC using at least three hits in both counters. The events were further selected by requiring the event vertex to be within ±12 cm of the VTPC center. From the selected events, those tracks with at least 11 hits in the VTPC and a good impact parameter were corrected for the geometrical and kinematical acceptance, tracking efficiency, charged particle background from particle decays and secondary interactions, were used in the determination of the pseudo-rapidity distribution. The tracking inefficiency due to the particles with $p_{\mathrm{T}} < 50$ MeV/c was estimated and corrected by extrapolating the fit to the transverse momentum distribution measurement at reference [57] to $p_{\mathrm{T}} = 0$. The detailed information about the analysis is available in [60].

Multiplicity distributions of charged particles

The CDF collaboration used the Run-1 data that has been used in the inclusive transverse momentum measurement also for measuring the charged particle multiplicity distributions at 630 GeV and 1800 GeV. The same selection criteria and corrections were also applied in this measurement.

Mean charged particle transverse momentum distribution

The average charged particle momentum per charged particle multiplicity was measured together with the inclusive transverse momentum measurements using the Run-1 and Run-2 data sets. The same event selection criteria and correction factors were applied to these measurements.

Transverse momentum distributions of e^+e^- pairs in Drell-Yan events

Using the data collected in 1992-1993 and 1994-1995 at 1800 GeV, the CDF collaboration made a measurement of the transverse momentum distributions of e^+e^- pairs in the range $66 - 116$ GeV$/c^2$, originating from resonant production and decays of Z^0 bosons. The event selection required the collision vertex to be within 60 cm of the nominal interaction position. The events with two or more electrons, at least one of them in the central region, were further filtered by requiring e^+e^- pair mass to be in the mass range 66 GeV$/c^2 \leq m_{e^+e^-} \leq$ 116 GeV$/c^2$. The selected pairs were used in the measurement of $\mathrm{d}\sigma\ /\mathrm{d}p_\mathrm{T}$. The details of the analysis are available in [61].

Underlying event in $Z \to e^+e^-$ events

The CDF Run-2 Drell-Yan events were also used for probing the underlying event. To ensure a good momentum resolution, the events were required to have only one vertex within 60 cm of the nominal interaction point. For e^+e^- events, at least one of the electrons was required to have $E_\mathrm{T} > 18$ GeV and in addition the electrons within $|\eta| < 1.1$ were required to have a track with $p_\mathrm{T} > 10$ GeV$/c$ coinciding with the calorimeter signal. For $\mu^+\mu^-$, events the muon tracks were required to have $p_\mathrm{T} > 18$ GeV$/c$, to lie in the range $|\eta| < 1$ and have a good χ^2 score from the track fitting algorithms. The pairs formed in both types of events were required to originate from within 4 cm of the established event vertex. Only the events where the lepton pair invariant mass was in the range 70 GeV$/c^2 < m_{l^+l^-} <$ 110 GeV$/c^2$ were used in the measurement of the charged particle multiplicity and transverse momentum distributions. The measurement were divided in three parts, called called toward, away and transverse. The "toward", "away" and "transverse" regions were defined in the transverse plane, for each event individually, with respect to the direction of the Z boson. The toward region was defined as the range $|\Delta\phi| < \pi/3$, transverse regions were $\pi/3 < |\Delta\phi| < 2\pi/3$ and away region was $2\pi/3 < |\Delta\phi|$ where $\Delta\phi$ is the azimuthal angle of the charged particle as measured from the Z-boson direction. Figure 5.3 shows the layout of the regions in $\Delta\phi$ plane. The transverse regions were labeled as Trans-min and Trans-max depending on the activity in the region. The details of the analysis and the data can be found in reference [62].

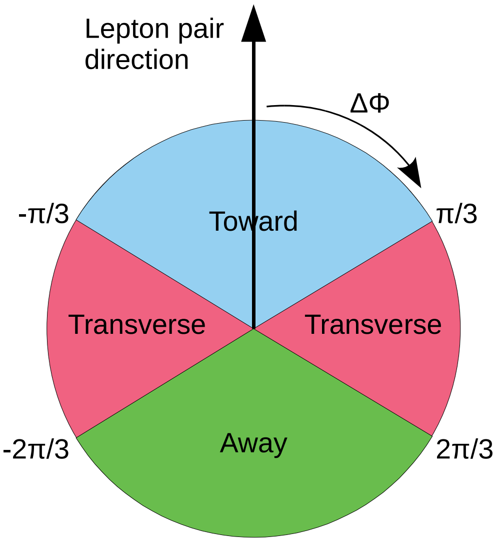


Figure 5.3.: Definition of towards, transverse and away regions used in the CDF underlying event analysis.

5.1.2. DØ

Like the CDF detector, the DØ detector is also located at the Tevatron collider at Fermilab. For the Run-2 phase of the Tevatron, the DØ detector went through some upgrades to cope with the higher data rates, enhance the physics reach of the experiment and also have a magnetic field to improve tracking [63]. The Silicon Microstrip Tracker (SMT) and a scintillating-fiber tracker, the CFT, inside a 2 T solenoidal superconducting magnet forms the central tracking system of the upgraded DØ detector. The cross-sectional view of the central tracking system can be seen in figure 5.4. Two tracking detectors can locate the primary interaction vertex with a resolution of about $35\,\mu$m along the beam line. At $|\eta| = 0$ they can tag b-quark jets with an impact parameter resolution of better than $15\,\mu$m in R–ϕ direction for the particles having $p_{\mathrm{T}} > 10$ GeV$/c$. With the help of the high resolution vertex positioning, the upgraded DØ detector achieves improved lepton p_{T}, jet-E_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement. Also the calibration of the electromagnetic calorimeters (EM) using the E/p ratio for the electrons became possible.

The calorimetric systems of the DØ detector cover the $|\eta| < 4$ region. The Central Calorimeter (CC) covers the range $|\eta| < 1$ and the remaining parts are covered by the two end-calorimeters. They use liquid argon as active material and have finer granularity towards the interaction region. The EM calorimeters are located closer to the center of the detector. There are pre-shower detectors in front of the EM calorimeter at the CC region for improved electron identification. The relative E_{T} resolution for the jets

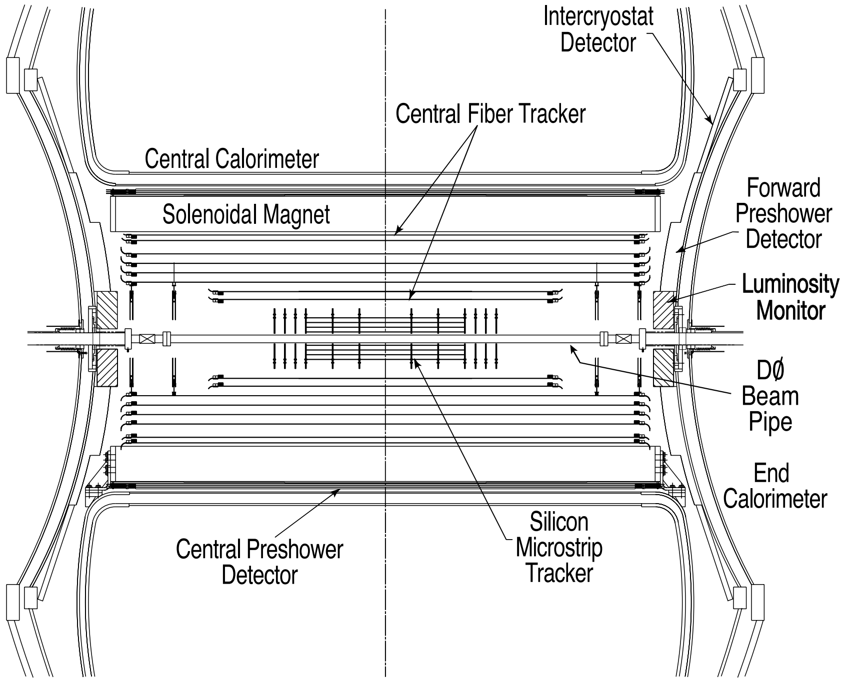


Figure 5.4.: Cross-sectional view of the central tracking system of the upgraded DØ detector.

reconstructed using a cone algorithm with $R = 0.7$ in $|\eta| < 0.4$ ranges from about 20% at 25 GeV to 6% at 330 GeV. Further details on the DØ detector can be found in [63] and the references therein.

Z Boson-Transverse momentum shape in Drell-Yan events

The DØ Collaboration made a measurement of the transverse momentum distribution of Z^0 bosons using the data collected during the Run-2 phase of the Tevatron. The data used in the analysis were selected by requiring two isolated electromagnetic clusters with shower shapes consistent with those of electrons. These electron candidates were required to have $p_{\mathrm{T}} > 25$ GeV/c and the invariant mass of the pair they form was required to be $70 < m_{e^+e^-} < 110$ GeV/c^2. In addition, the central region electrons were required to have a matching track in the trackers. The successful events were corrected for the efficiencies and acceptances and used in the measurement of $1/\sigma d\sigma/dp_{\mathrm{T}}$ of the Z^0 bosons in the ranges $40 < m_{Z^0} < 200$ GeV/c^2 and $p_{\mathrm{T}} < 260$ GeV/c. The Z^0 bosons produced at rapidities larger than 2 were also used in the measurement of $1/\sigma\, \mathrm{d}\sigma\, /\mathrm{d}p_{\mathrm{T}}$ for the domain $p_{\mathrm{T}} < 30$ GeV/c in order to test the small-x predictions of the theories. The specifics of the analysis is given in [64].

5.1.3. UA5

The UA5 detector was one the experiments at the CERN SPPS proton anti-proton collider. It was a non-magnetic detector and was optimized for studying the charged particle multiplicities in non-single-diffractive events. In the UA5 detector, the large streamer chambers placed above and below the beryllium beam pipe provided about 95% geometric acceptance in the $|\eta| < 3.0$ range, falling to zero at $|\eta| < 5.0$. The scintillator counter hodoscopes, covering the pseudorapidity ranges $2.0 < |\eta| < 5.6$ were used for triggering the high voltage pulse to the streamer chambers and recording the event on photographic films. The trigger logic required at least one hit in both trigger hodoscopes. The events were recorded onto the photographic films by three sets of cameras each with a pair of stereo views directed towards the up and down streamer chambers. Figure 5.5 shows a schematic layout of the UA5 detector. Reference [65] contains an extensive description of the UA5 experiment.

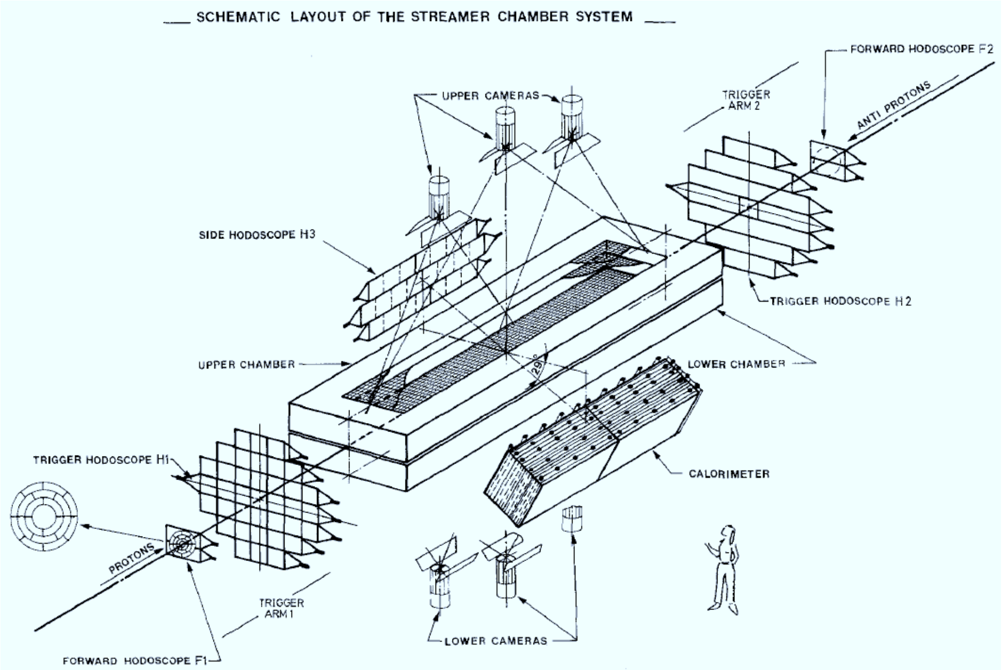


Figure 5.5.: Schematic view of the UA5 Detector.

Multiplicity distributions of charged particles

The UA5 collaboration made several measurements of multiplicity distributions at center-of-mass energies of 200 GeV, 546 GeV and 900 GeV. The tracks contained in the data acquired at 200 GeV and 900 GeV were reconstructed by fitting the track images to a straight line. After all the tracks in an event were reconstructed, the event vertex was determined. The events were corrected for the efficiencies and acceptances using unfolding by maximizing the entropy of the distributions of the triggered events. The details of the procedure is described in [66]. The corrected events were used in the measurements of the charged particle multiplicity distributions in the full detector acceptance range as well as in the symmetric pseudorapidity intervals of $|\eta| < 0.5, 1.5, 3.0$ and 5.0. For the data at 546 GeV, a similar procedure was used in event reconstruction, however the efficiency and acceptance corrections were done with another method which introduced artificial oscillations in the measurements. Details of the method is available in [65] and reasons of oscillations are discussed in [66].

Pseudorapidity distributions of charged particles

The pseudorapidity distribution measurements by the UA5 collaboration were done by using the one-arm or two-arm triggers, corresponding to at least one hit on the scintillator counter hodoscopes on either one or both sides of the detector in the range $2.0 < |\eta| < 5.6$, respectively. The two-arm trigger was sensitive mainly to the non-single-diffractive events, on the other hand the one-arm trigger selected highly asymmetric events such as the single-diffractive events. The events were corrected for secondary particle decays, efficiencies and acceptances. The $\mathrm{d}n\,/\mathrm{d}\eta$ measurements were done using the non-single-diffractive and inelastic events collected at 200 GeV, 546 GeV and 900 GeV in $|\eta| < 5.0$. The $\mathrm{d}n\,/\mathrm{d}\eta$ distributions at different charged particle multiplicity intervals were also measured using the same data sets. The details of the analysis are presented in [65, 67].

5.1.4. Implementation of analyses

MC generators usually have different event generation styles and definitions of event types such as diffractive and non-diffractive. Thus to compare the data sets generated by the MC generators and data, one must put them on equal footing as much as possible, in a way as independent from generator specific definitions or styles as possible. Therefore a library creating the distributions from MC events has been implemented. For all data sets, the bins of generated sets are arranged such that either the data point is in the center of the bin or the same as the data binning whenever available. The library includes basic imitation of the triggers used in experiments for the event selection whenever possible to use the event definitions of the experiments rather than MC.

The UA5 data sets used the requirement of at least one charged particle in the range $2.0 < |\eta| < 5.6$ in both directions as trigger.

For the CDF minimum bias data sets, the Run-1 trigger implementation required at least one charged particle in both directions $3.2 < |\eta| < 5.9$. The Run-0 trigger required a minimum of four charged particles in the range $|\eta| < 3.0$, at least one in both $+\eta$ and $-\eta$ hemispheres, on top of the Run-1 trigger requirements. The Run-2 trigger implementation required at least one charged particle in both directions in the range $3.7 < |\eta| < 4.7$. For the underlying event data sets, the trigger implementation required at least two electrons within the range $-1.0 < \eta < 1.0$ and with $p_\mathrm{T} > 20$ GeV/c with the hardest pair satisfying the cuts $70 < m_\mathrm{pair} < 100$ GeV/c^2 and $|\eta_\mathrm{pair}| < 6.0$.

For simplicity, the DØ Drell-Yan Z-boson transverse momentum shape measurement required only the existence of a Z boson with the mass in the range $40 < m_{Z^0} < 200\ \mathrm{GeV}/c^2$.

5.2. Genetic Algorithms

The Genetic Algorithms (GA) belong to a class of algorithms that are called Evolutionary Algorithms which are derived from our observations of the nature and Evolution Principle [68]. A GA tries to mimic the evolution of a species by employing the principle of “survival of the fittest” for searching an approximate or exact solution of optimization or search problems. In the nature an individual goes through the trials of life. A successful individual usually ends up with more essential resources such as the food, space or mating chance. Thus it is more likely to have more descendants where it passes its genes on. The GA’s work in the same way, the individuals are tested for the fitness by a scoring function, and the successful candidates are given a chance to procreate for the next generation.

The individuals in a GA represent a solution candidate for the problem. They are one of the two most important components of a GA and can be in very different types depending on the problem at hand. They have one or more parameters which are called genes. In the most basic case, a gene is represented by a bit-string, i.e a series of 0’s and 1’s, like the DNA of a being which has only two amino acids. It is possible to map this bit-string to the other types of the data such as an integer, where each bit in the bit-string corresponds to a bit of the integer or a real number where it represents the discrete values within a given range. Nowadays it is possible to find almost all basic data types implemented in publicly available GA libraries.

The second most important component of a GA is the “fitness function” also called as “evaluation function” or “scoring function”, which evaluates an individual against the problem and assigns a measure to determine its chance of survival and procreate. The scoring function is the representation of the problem, thus the better the implementation of the problem as a function, the better the results get.

The GA’s have two main operations, namely the “Cross-over” and “Mutation” for generating new individuals out of the old ones. The cross-over operation represents the procreation of individuals and defines the rules of how a descendant is created from one or more parents. The cross-over can be among multiple individuals (sexual procreation) or on a single individual (asexual procreation). A basic cross-over operation can be seen in figure 5.6. The Mutation, on the other hand, alters one or more genes of the individuals changing their location in the solution space. Figure 5.7 shows an example of a sexual procreation process for two parents who have two genomes representing the x and y coordinates in the x-y plane. The cross-over operation creates the child, at a point on the line segment between the mother and father. Notice that it also includes some extra segments near parents so that a child could be similar to one of the parents and dissimilar from the other. After the child is created by the cross-over, the mutation, by a given probability, shifts the point randomly within a circle of pre-defined radius. Usually the cross-over operation concentrates the individuals around a possible solution and the mutation helps escaping from the local extrema conditions.

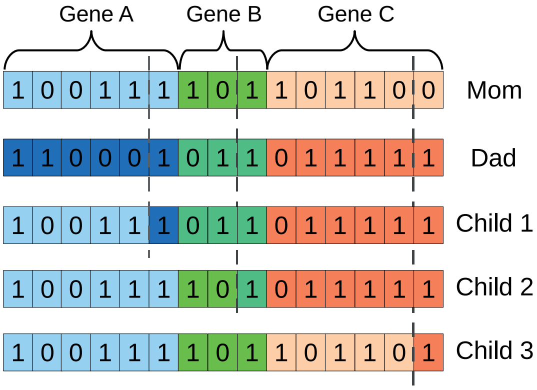


Figure 5.6.: An example of procreation using a basic cross-over operation. The parents have three genes and the children are generated by selecting a random point for cutting the genes. The right-hand side of cutting point comes from the dad and the left-hand side comes from the mom.

The collection of individuals in a GA is called population. The GA evaluates the fitness of all individuals, then does a pairing among them applying the cross-over and mutation operations to the pairs to create the new individuals. Usually the individuals with higher scores have higher changes to procreate. The evaluation-procreation cycle is repeated until the desired convergence or termination conditions are satisfied. The population at each cycle is called a generation. The population sizes can be constant or variable throughout the lifetime of the program.

A basic demonstration of a GA is shown in figure 5.8. The red curves in the plots show the function

$$f(x) = \frac{x^2}{25.0} + 0.005 + (0.01\mathrm{R}(0, 0.005))\sin(x(5 + \mathrm{R}(0, 0.05))) + 1$$

where $\mathrm{R}(0, 0.05)$ is a normally distributed random number, with a mean of 0 and a standard deviation (σ) of 0.05, evaluated at every x to create a small random perturbation. The green curve is a singular value decomposition parametrization of a second order polynomial using 300 random points on the function. The blue dots show the scores of the individuals in GA. The top left plot shows the randomly initialized individuals in the GA. The top right plot and the plots in the middle rows show the evolution of the individuals at generations one to five. The bottom row contains the plots of the individuals at the 10^{th} and 30^{th} generations. As can be seen from the plots, the GA converges to the absolute minimum of the function pretty quickly and does a better job than the singular value decomposition solution.

Genetic Algorithms are a very diverse subject and there are lots of different types of implementation. For more detailed discussions see [69–71] and references therein.

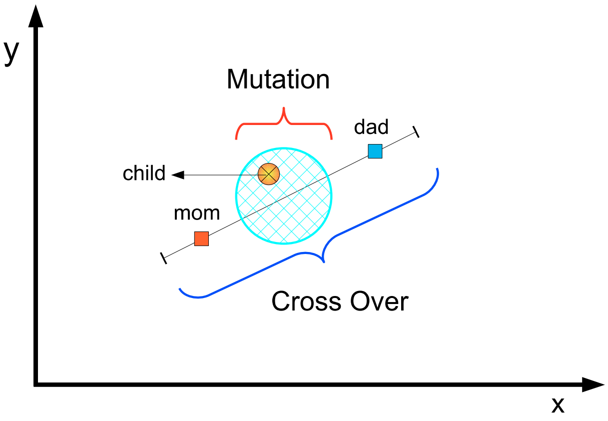


Figure 5.7.: Generation of a new individual from two parents. The First a point between the mom and dad chosen randomly by the cross-over and then altered by the mutation.

5.3. Application of GA's to MC Tuning

As mentioned in the Chapter 4 some MC programs have user tunable parameters. In the tuning process, large MC samples are generated where the detector acceptances and efficiencies as described in respective articles given in section 5.1 are applied. Then these distributions are compared to the measurements. The program parameters are altered depending on the results of the comparison and the same steps are repeated.

For a small number of parameters and data sets, such a task can be done manually. However as the number of parameters or data sets increase, the number of histograms to check increase exponentially. For example, for a 5 parameter tuning with 5 data sets and taking 10 different points for each parameter one must check $5 \text{ x } 10^5 = 500000$ histograms. Obviously this can not be done manually and an automatisation is needed. However the generator response is a complex function of input parameters although bin by bin parametrization of the generator response can help in automatization, it approximates the response domain.

The large volume of the parameter space that is usually encountered in MC tuning, makes it harder to find the optimum parameters. As described in the previous section, GA's have good applications in search problems. By using a goodness-of-fit (GOF) function between the data and MC distributions and assigning the parameters as the genomes of individuals, it is possible to use the GA's for searching for the optimum parameter sets using the exact generator response.

For exploiting the power of the GA in the MC tuning an analysis framework called GAMPI Tuner has been prepared. GAMPI is an acronym for the **G**enetic **A**lgorithm with **M**essage **P**assing **I**nterface. The GAMPI is composed of several parts:

- Generator Module

- Analysis Module
- Genetic Algorithm Module
- Distribution Module

GAMPI is implemented in C++ and has an interface for the FORTRAN MC generators. It is broken into modules, each with a different purpose, and a schematic organization of the modules can be seen in figure 5.9. This modular structure makes it easy to extend it to other generators and analyses.

5.3.1. Generator Module

The generator module of GAMPI is responsible for the event generation. It is linked against the generator library. It receives the parameters from the Distribution Module, initializes the MC event generator, generates the events and passes the event information to the Analysis Module. This module can be re-implemented or the two existing implementations using PYTHIA 8 and PYTHIA 6 can easily be adopted for the other MC event generators.

5.3.2. Analysis Module

The analysis module contains the analysis code that creates the distributions, the functions that define the fitness of distributions with respect to the data and the data itself. This module corresponds to the evaluation function of the GA, thus an accurate implementation improves the results. The existing analysis library contains the analysis code for the data described in section 5.1. It can be extended or interfaced with another analysis library such as the HzTools [72] or Rivet [48–50].

The analyses with common beam types, collision energy and event types can share the events generated by the MC program, however in most of the time it is necessary to generate separate event sets because of differences in data taking conditions. Thus the evaluation time of a parameter set is usually proportional to the number of analyses it has and the number of events generated. For the analyses used in this thesis, the evaluation of a parameter set with one million events took between three to six hours on a 3 GHz Xeon CPU depending on the generator and the parameters.

5.3.3. Genetic Algorithm Module

The Genetic Algorithm module is implemented using a publicly available library called GAlib [69]. It is a collection of C++ classes and has basic genetic structures and operations defined. In this thesis it is extended to suit the needs of the MC tuning problem.

In the current implementation of the module, the individuals are composed of real valued genes that can take any value within the bounds up to the floating point precision. For almost all of the generators the real precision is more than enough since tiny differences in parameters do not change the output of the generator. The individuals have many genes, each corresponding to a parameter of the MC. They are assigned a mating chance according to the Roulette Wheel schema, that is the higher scoring individuals are given a higher mating chance proportional to their scores.

For simplicity a constant-sized population model called "Steady State" population has been chosen. In this model the population size does not change, however, for each generation a given portion of the worst scoring individuals are replaced by the best scoring individuals generated from the current generation. Generation of the new individuals are done by sexual procreation. The parents are selected randomly depending on their mating chances, then each pair is tried against a given cross-over chance. The accepted pairs go through the cross-over operation.

The cross-over operation creates the children from the parents by selecting a random point for each gene in the range

$$\left[g_n^{\mathrm{dad}} - 0.5(g_n^{\mathrm{mom}} - g_n^{\mathrm{dad}}), g_n^{\mathrm{mom}} + 0.5(g_n^{\mathrm{mom}} - g_n^{\mathrm{dad}})\right],$$

where g_n^{mom} and g_n^{dad} are the n^{th} gene of the mom and dad respectively. The extra segments at the edges of the range means that the gene of the child has 50% change to lie somewhere in between the mom and dad and 25% chance to lie away from both but relatively closer to one of them.

After a child is created, it is tested whether it undergoes mutation. The mutation operator randomly selects a gene of the child and shifts it randomly according to a normal distribution. The width of the normal distribution is a parameter of the program and defined in terms of the bounds of the given gene. That is

$$g_{\mathrm{mut}} = g_{\mathrm{orig}} \frac{\mathcal{N}}{\mathrm{w}} (g_{\mathrm{u}} - g_{\mathrm{l}})$$

where the g_{mut} is the mutated value, g_{orig} is the original value, $\mathcal{N}$ is the normally distributed random number, w is the width parameter and the g_{u} and g_{l} are the upper and lower bounds of the gene coming from the boundaries of the parameter space.

5.3.4. Distribution Module

The distribution module is composed of two parts, the Master process and the Slave processes. The Master process is interfaced with the genetic algorithm and handles the communication between GA and the computers where the slave processes run. The Slave process on the other hand drives the generator module and the analysis module, receives the parameters from the Master process and sends the results back. Both modules are implemented using OpenMPI [73]. Since OpenMPI is compliant with the MPI 2.0 standard, both modules can be used with other MPI implementations with minimal modifications.

The Master process also handles the parallelization of the tuning problem. Since the analysis module is composed of several analyses, which require distinct settings for the same parameters, the master process sends each analysis to a different slave, efficiently using the available resources. It creates a small data packet containing the parameters, analysis type, number of events to be generated, collision energy, timeout length and the individual number. For the current implementation this package can contain up to 20 parameters from the individuals. The package is distributed to the slave processes through the MPI.

The Slave process initializes the generator module for the given analysis mode with the given collision energy and parameters. It drives the event generation and checks for the processes that are too slow at generating events. Checking for such processes conserves a lot of CPU time since slow generation usually happens due to a bad choice of parameters which typically lead to too small kinematical domains or at the edges of some parameter domains where certain assumptions made by the MC authors become invalid. When this happens the Slave process aborts the event generation and assigns a very bad score for the parameter set and analysis. Finally it gets the scores of the analysis from the analysis library and sends them back together with the total processing duration.

The distribution module uses the process pool approach. When the processes are started, Slave processes ask the master process for the parameters and the master process replies with the parameters from the GA module. When the Slave process completes the analysis, it sends the results back, asking for new sets. When the Master process receives all results of a generation, it combines the results and transfers the scores to the GA, repeating the process for the next generation.

5.4. Tuning PYTHIA 6 and PYTHIA 8

For some data sets it is not clearly explained how the data were collected, how corrections were applied or later it is found that the analysis method introduced artifacts to data, such as the artificial oscillations observed in the distributions as explained in [66]. Because of these reasons only a subset of the data described in section 5.1 is used in the tuning process. Table 5.1 shows the data sets used in the the tuning process. The CDF Lorentz invariant cross-section measurements are converted to normalized p_T distributions through the formula

$$\frac{\mathrm{d}N_{ch}}{\mathrm{d}p_\mathrm{T}}[i] = \frac{p_\mathrm{T}[i]\cdot \mathrm{Linv}[i]}{\sum_{j=\mathrm{first}}^{\mathrm{last}} \mathrm{Linv}[j]\cdot p_\mathrm{T}[j]}$$

where $[i]$ and $[j]$ represents the i-th and j-th bins and Linv is the Lorentz invariant cross-section $E\frac{\mathrm{d}^3\sigma}{\mathrm{d}p_\mathrm{T}^3}$. Drell-Yan measurements are also normalized to one since the leading order MCs can not describe the cross-section at a precision of experimental measurements.

Seven parameters in PYTHIA 8 and ten parameters in PYTHIA 6 are tuned. Since this study is aimed for minimum bias analysis with some underlying event component, only a subset of parameters can be constrained with the data used for tuning. Thus some parameters such as fragmentation and hadronization parameters tuned to LEP data are taken from [52] and listed in table 5.4. The default fragmentation and hadronization parameters are not changed for PYTHIA 8. Table 5.2 lists the parameters tuned in PYTHIA 8 and table 5.3 lists for the PYTHIA 6.

Coll.	$\sqrt{s}$	**Data Set**	**Ref.**
CDF	1960 GeV	$< p_{\mathrm{T}} >$	[59]
		dN_{ch}/dp_{T}	
		UE Trans-Min $\sum p_{\mathrm{T}}$ density in Drell-Yan	[62]
		UE Trans-Max $\sum p_{\mathrm{T}}$ density in Drell-Yan	
		UE Trans-Min $\sum N_{ch}$ density in Drell-Yan	
		UE Trans-Max $\sum N_{ch}$ density in Drell-Yan	
	1800 GeV	dN_{ch}/dp_{T}	[57]
		$1/\sigma d\sigma/dp_{\mathrm{T}}$ (Drell-Yan)	[60]
	630 GeV	dN_{ch}/dp_{T}	[60]
DØ	1960 GeV	$d\sigma/dp_{\mathrm{T}}$ Drell Yan	[64]
		$d\sigma/dp_{\mathrm{T}}$, $y > 2$ $p_{\mathrm{T}} < 30$ GeV	
UA5	900 GeV	$dN_{ch}/d\eta$	[67]
		N_{ch}	[66]
		N_{ch}, $\lvert\eta\rvert < 0.5$	
		N_{ch}, $\lvert\eta\rvert < 1.5$	
		N_{ch}, $\lvert\eta\rvert < 3.0$	
		N_{ch}, $\lvert\eta\rvert < 5.0$	
	546 GeV	N_{ch}	[65]
	200 GeV	$dN_{ch}/d\eta$	[67]
		N_{ch}	[66]
		N_{ch}, $\lvert\eta\rvert < 0.5$	
		N_{ch}, $\lvert\eta\rvert < 1.5$	
		N_{ch}, $\lvert\eta\rvert < 3.0$	

Table 5.1.: Data sets used in tuning process. The details of the analyses are explained in section 5.1 and in their respective references.

Due to the nature of the models, it is very hard to directly relate the parameters to the data. However some loose relations can be made. The average p_{T} distribution is related to the number of multiple parton interactions and color flow. Charged particle distributions are related to the multiple parton interactions and matter distribution and Drell-Yan distributions can be related to coupling constant. Transverse regions in underlying events are sensitive to the beam remnants.

Three different goodness of fit functions are used in tuning. These are

$$\mathrm{Score}_0(H, D) = \sum_{i=\mathrm{Firstbin}}^{\mathrm{Lastbin}} \frac{(D_i - H_i)^2}{D_i^2} \tag{5.1}$$

$$\mathrm{Score}_1(H, D) = \sum_{i=\mathrm{Firstbin}}^{\mathrm{Lastbin}} \frac{(D_i - H_i)^2}{(\mathrm{Avg}(D) \cdot D_i)} \tag{5.2}$$

$$\mathrm{Score}_2(H, D) = \sum_{i=\mathrm{Firstbin}}^{\mathrm{Lastbin}} \frac{(D_i - H_i)^2}{(\sigma_{\mathrm{D_i}})^2} \tag{5.3}$$

where H_i and D_i are the i-th bin of the produced histogram and the data respectively. The function Avg(D) gives the average value of the data between the Firstbin and the

Parameter	**Explanation**
MultipleInteractions:pT0Ref	$p_{\mathrm{T}_0}^{ref}$ in equation 4.28. Used in the regularization of the hard cross-section divergence.
MultipleInteractions:ecmPow	E^{pow} in equations 4.27 and 4.28. It is also used in the regularization of the hard cross-section divergence.
MultipleInteractions:coreRadius	gives the relation between a_1 and a_2 in equation 4.24 such that the inner radius is assumed to be smaller by this factor than the outer radius.
MultipleInteractions:coreFraction	β factor in equation 4.24. It controls the fraction of the matter in inner core and outer core.
BeamRemnants:primordialKThard	σ_{hard} in equation 4.30. It controls the width of the Gaussian distribution used in determination of the primordial kT.
BeamRemnants:reconnectRange	Controls the probability of merging the gluons in subprocesses in the multiple interactions.
MultipleInteractions:alphaSvalue	The value of the strong coupling constant at Z^0 mass, $\alpha_s(m_{Z^0})$.

Table 5.2.: List of tuned parameters in PYTHIA 8. Detailed explanations are available in [36].

Lastbin and $\sigma_{\mathrm{D_i}}$ is the error of the i-th bin of the data. The first function Score$_0$ is sensitive to the variations at the tails of the distribution and the second one, Score$_1$, is a compromise between the tails and the peaks with a bias towards the peaks. The third function, Score$_2$, is the generic χ^2 function. Since the treatment and reporting of the errors in the data are inconsistent among data-sets, χ^2 is not entirely reliable. Considering that it is almost impossible to find a perfect description of the data by MC everywhere, these three functions provide a choice for the user for optimizing the desired parts of the distributions. The final fitness score of a parameter set is found by adding the scores of the distributions for all data-sets normalized with the number of bins used. In this thesis, all scoring functions are used in the tuning process, however the function Score$_1$ yielded best results for the overall description.

Possibly the most important property of GAMPI is its ability to distribute the event generation to several computers, since an increased number of individuals and statistics increases the scanning efficiency. In order to make use of this ability, GAMPI is run on 160 CPUs, generating one million events per minimum bias and 250 thousand events per Drell-Yan data sets described in section 5.1. Different population sizes, scoring functions, cross-over and mutation chances are tried. At each generation half the size of the population, N/2 new individuals are generated and best N individuals of all are

Parameter	**Explanation**
PARP[64]	Controls the value of α_s scale in the space-like showers such that $\alpha_s(\mathrm{PARP}[64] \cdot k_{\mathrm{T}}^2)$.
PARP[71]	Controls the maximum parton virtuality allowed in time-like showers.
PARP[78]	Controls the color reconnections in the final state.
PARP[79]	Controls the x values of the Beam Remnants.
PARP[82]	p_{T_0} in equation 4.28
PARP[83]	d in equation 4.25
PARP[90]	E^{pow} in equations 4.27 and 4.28
PARP[91]	σ and $\sqrt{< k_{\mathrm{T}}^2 >}$ in equation 4.29
PARP[93]	Upper cut-off for the primordial k_{T} given in equation 4.29
PARP[101]	Controls the fraction of the two different diffractive event structures.

Table 5.3.: List of tuned parameters in PYTHIA 6. Detailed explanations are available in [32].

Parameter	**Value**	**Parameter**	**Value**	**Parameter**	**Value**
MSTP[41]	12.	PARP[80]	0.01	PARP[89]	1800
MSTP[51]	7.0	MSTP[52]	1.0	MSTP[68]	3.0
MSTP[70]	2.0	MSTP[72]	0.0	MSTP[81]	21.0
MSTP[82]	5.0	MSTP[88]	0.0	MSTP[89]	1.0
MSTP[91]	1.0	MSTP[95]	6.0	PARJ[1]	0.073
PARJ[2]	0.2	PARJ[3]	0.94	PARJ[4]	0.032
PARJ[11]	0.31	PARJ[12]	0.4	PARJ[13]	0.54
PARJ[25]	0.63	PARJ[26]	0.12	MSTJ[11]	5.
PARJ[21]	0.313	PARJ[41]	0.49	PARJ[42]	1.2
PARJ[47]	1.0	PARJ[81]	0.257	PARJ[82]	0.8
MSTJ[21]	1.				

Table 5.4.: Parameter choices for PYTHIA 6. Composed of fragmentation and flavor parameters tuned by the Professor tool using the LEP data [52].

allowed to progress to the next generation.

Some example plots showing the results of tuning using two different diffraction models in PYTHIA 8 and the default values are shown in figures 5.10 to 5.14. Significant improvements in description of data are clearly visible.A comparison of the results with the other tunes is presented in the next section.

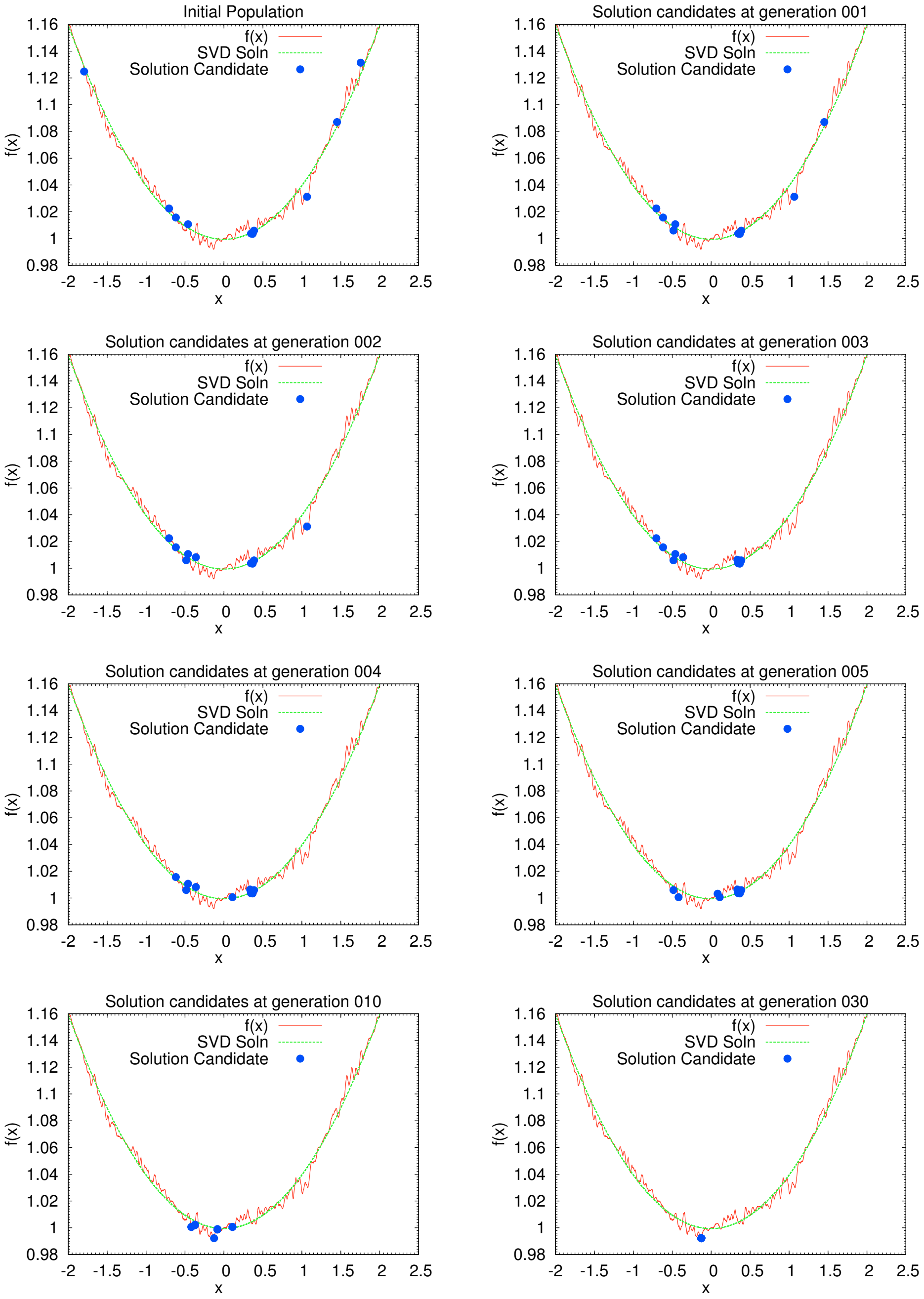


Figure 5.8.: Evolution of the solution of a function with a random sinusoidal noise component. The top left plot is the scores of the randomly generated initial solution candidates. Top right and the center rows show the candidates after the generations 1 to 5 respectively. The bottom row shows the results after the 10th and 30th generations respectively. The SVD parametrization is also shown in the plots.

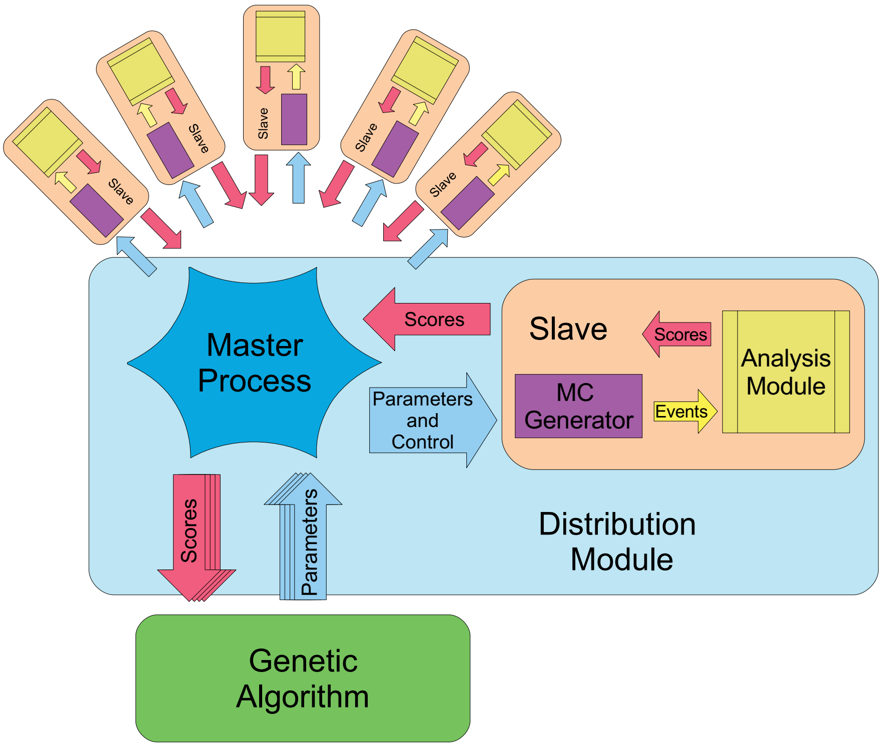


Figure 5.9.: Schematic layout of GAMPI. Different components of the program are seen in the figure. MC Generator and Analysis modules form the Slaves. Slaves together with the Master Process form the Distribution module. Master process controls several slave processes. Genetic Algorithm module talks with the Master Process to pass on parameters and get the scores back.

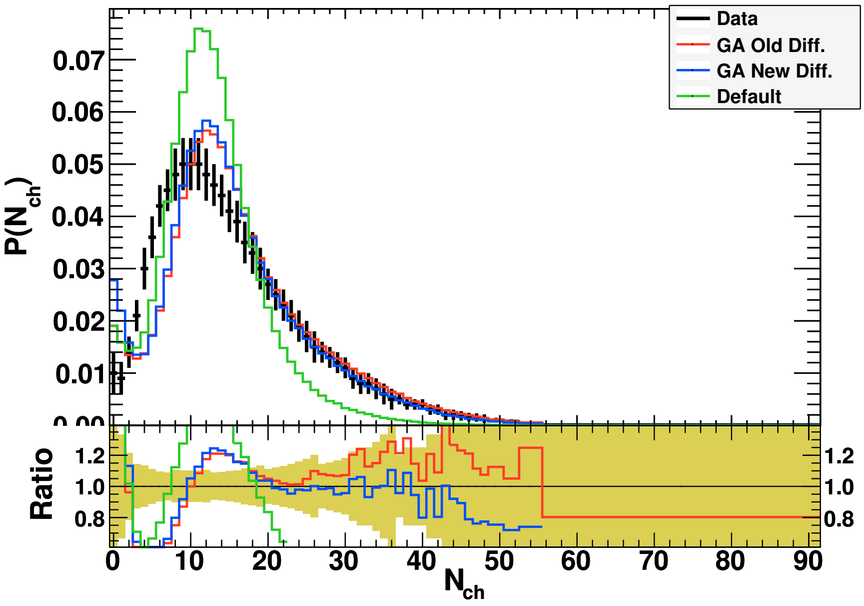


Figure 5.10.: Charged particle multiplicity distributions from UA5 at $\sqrt{s} = 200$ GeV in $|\eta| < 3$.

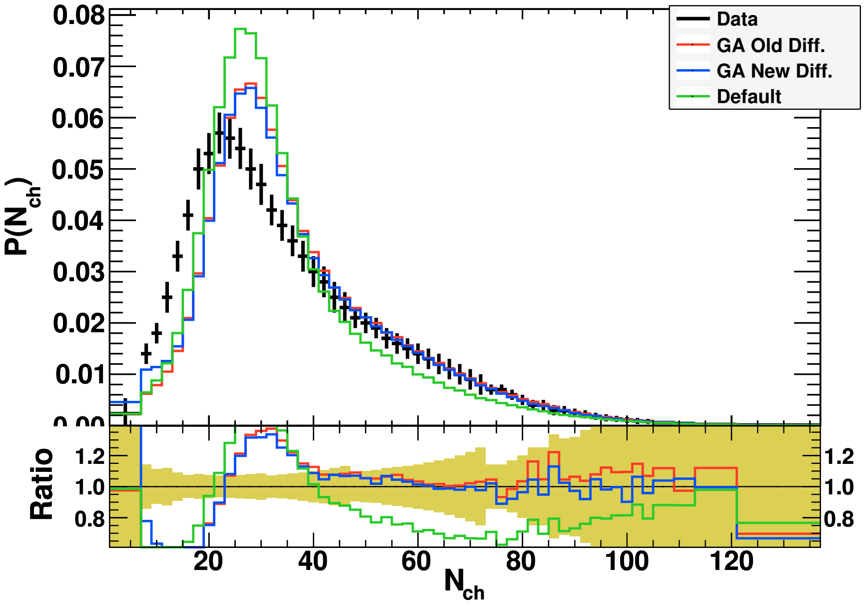


Figure 5.11.: Charged particle multiplicity distributions from UA5 at $\sqrt{s} = 900$ GeV in $|\eta| < 5$.

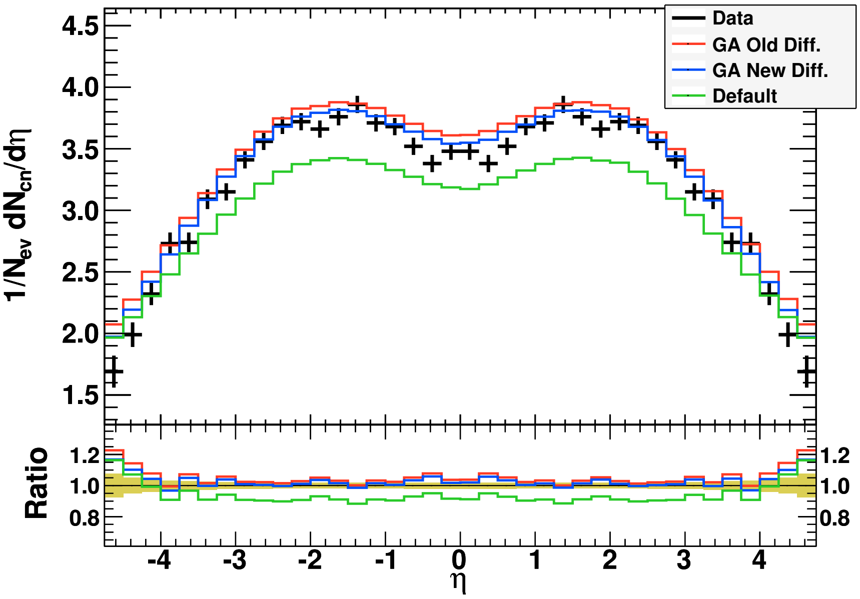


Figure 5.12.: Charged particle pseudorapidity distributions from UA5 at $\sqrt{s} = 900$ GeV.

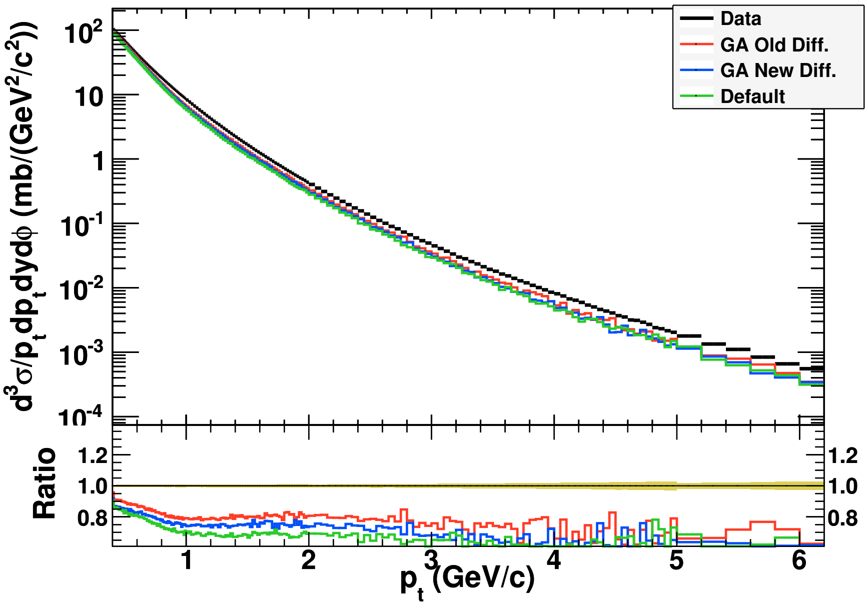


Figure 5.13.: Charged particle p_{T} distributions from CDF at $\sqrt{s} = 1960$ GeV in $|\eta| < 1$.

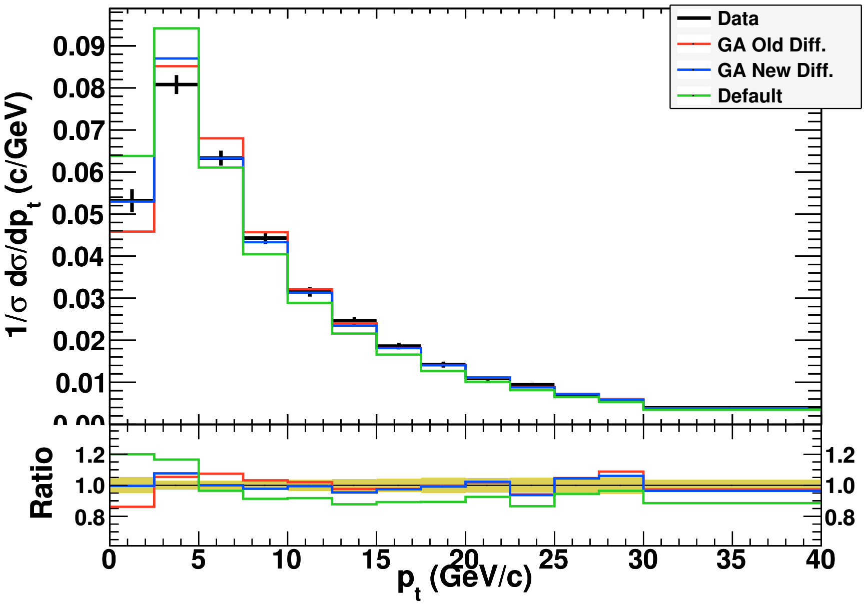


Figure 5.14.: Z boson p_{T} distribution in Drell-Yan events from DØ at $\sqrt{s} = 1960$ GeV.

5.5. Results

Tuning with the GA algorithm produces several different sets, usually improving as the generations progress. The tunes given in tables 5.5 and 5.6 for PYTHIA 8 and PYTHIA 6 are compared to the tune found by Professor tool [52] for PYTHIA 6 and the default tune of PYTHIA 8 starting from the version 8.130.

As can be seen from figures 5.15 to 5.24, GA tunes are comparable to or better than the alternative tunes. A similar behavior is observable in other plots of the data sets. In all of the plots black crosses represent the data, red and green curves are results of the GA tune for PYTHIA 8 and PYTHIA 6, respectively. The blue curves are the default tune for PYTHIA 8 version 8.130 and the magenta curves are the Professor tune for PYTHIA 6. The yellow zones in the ratio plots are the errors on data. Appendix C contains comparisons with several other tunes.

Parameter	**Value**
MultipleInteractions:pT0Ref	2.018883
MultipleInteractions:ecmPow	0.2468373
MultipleInteractions:coreRadius	0.57937294
MultipleInteractions:coreFraction	0.26321179
BeamRemnants:primordialKThard	1.8995764
BeamRemnants:reconnectRange	6.5093927
MultipleInteractions:alphaSvalue	0.1193707

MultipleInteractions:pTmin	0.2
MultipleInteractions:bProfile	2
MultipleInteractions:ecmRef	1960.
Tune:ee	3
BeamRemnants:valenceDiqEnhance	5
Diffraction:PomFlux	3

Table 5.5.: GA Tune results for PYTHIA 8. Bottom part of the table shows the parameters for flavor, fragmentation, hard p_{T} cutoff, its energy dependence pivot point and new diffractive scheme setups.

Parameter	**Value**
PARP[64]	0.5463903
PARP[78]	0.15877976
PARP[82]	1.9125689
PARP[90]	0.21483184
PARP[93]	5.5663948

Parameter	**Value**
PARP[71]	3.0709131
PARP[79]	2.0780785
PARP[83]	2.01578
PARP[91]	2.0690465
PARP[101]	0.29185629

Table 5.6.: GA Tune results for PYTHIA 6. In addition to these values, the values in table 5.4 should be used for their respective parameters.

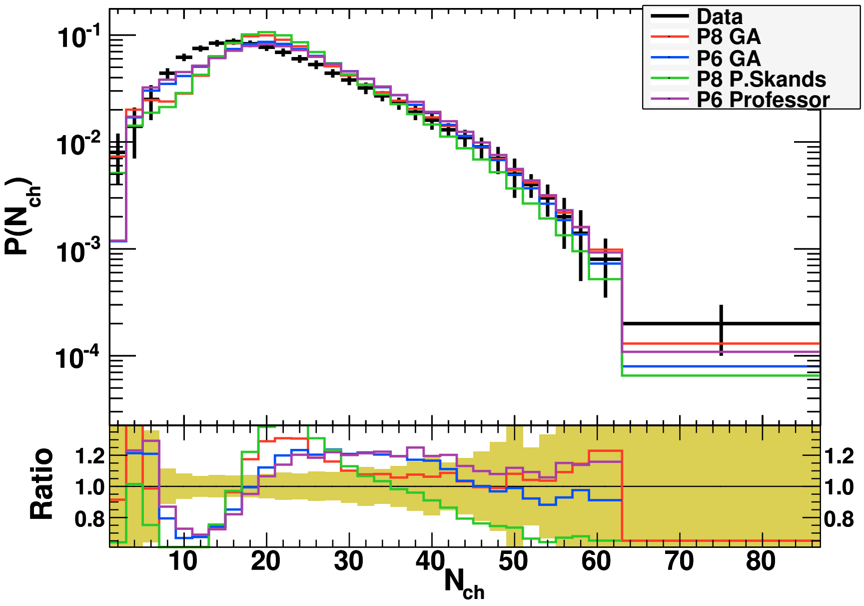


Figure 5.15.: Charged particle multiplicity distributions from UA5 at $\sqrt{s} = 200$ GeV in $|\eta| < 5$.

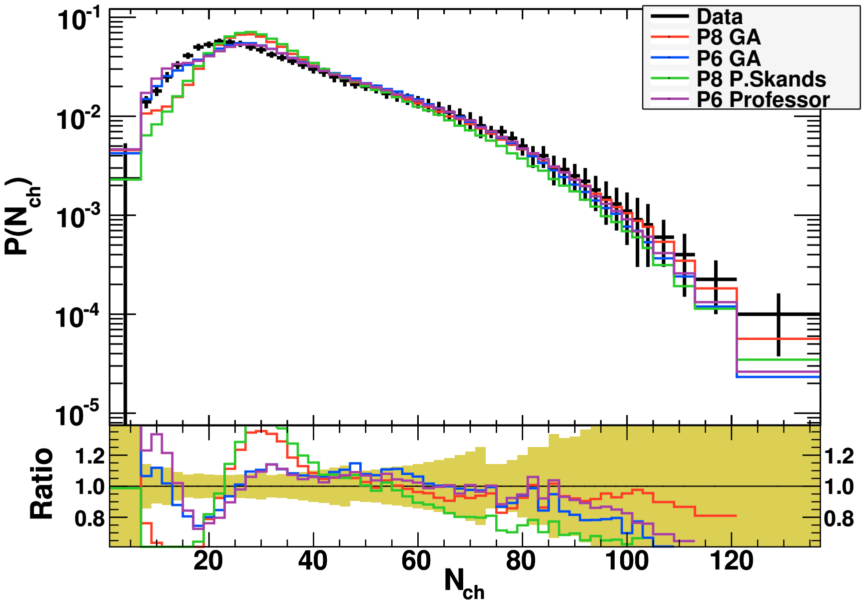


Figure 5.16.: Charged particle multiplicity distributions from UA5 at $\sqrt{s} = 900$ GeV in $|\eta| < 5$.

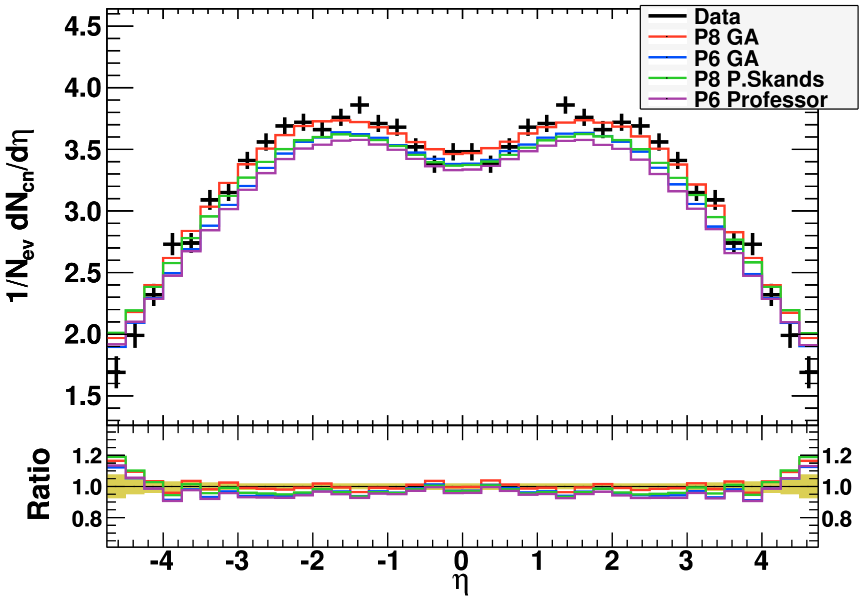


Figure 5.17.: Charged particle pseudorapidity distributions from UA5 at $\sqrt{s} = 900$ GeV.

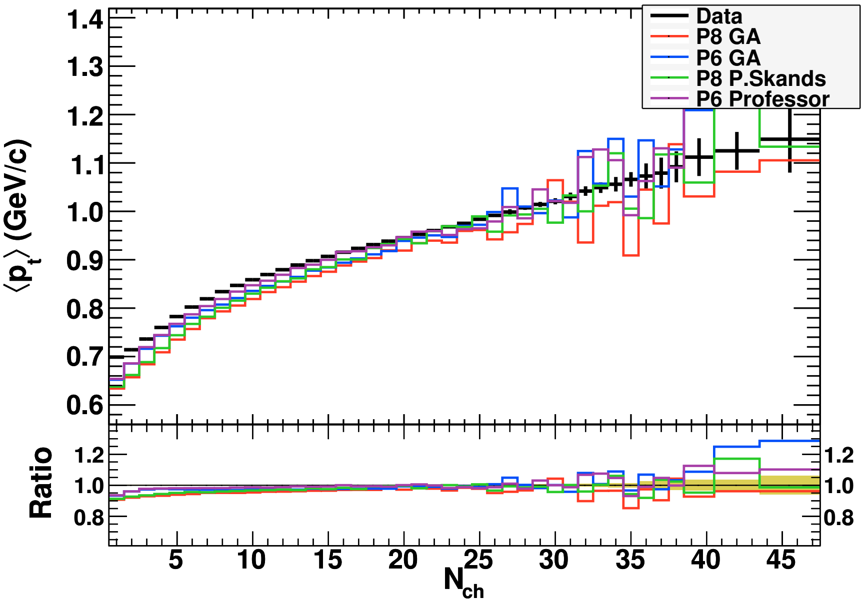


Figure 5.18.: Average charged particle momentum distributions from CDF at $\sqrt{s} = 1960$ GeV in $|\eta| < 1$.

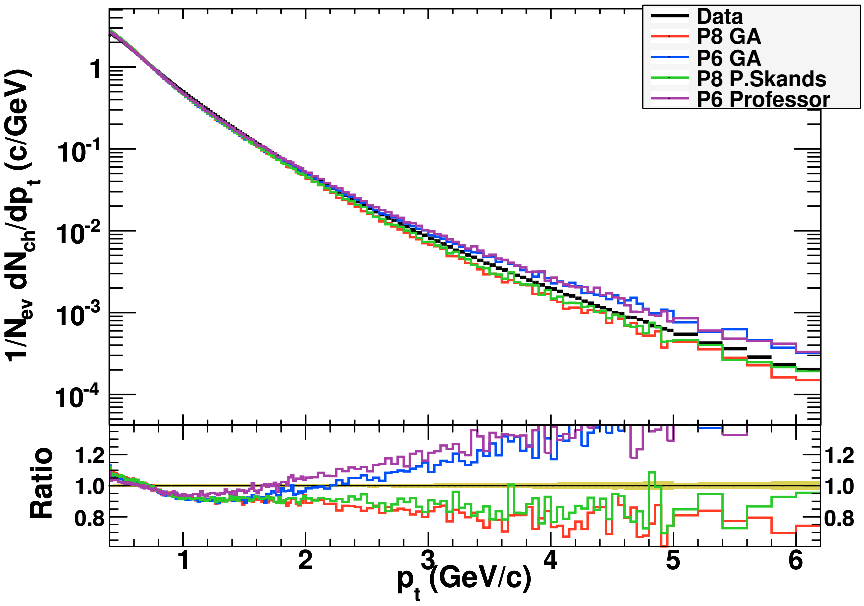


Figure 5.19.: Charged particle transverse momentum distributions from CDF at $\sqrt{s} = 1960$ GeV in $|\eta| < 1$.

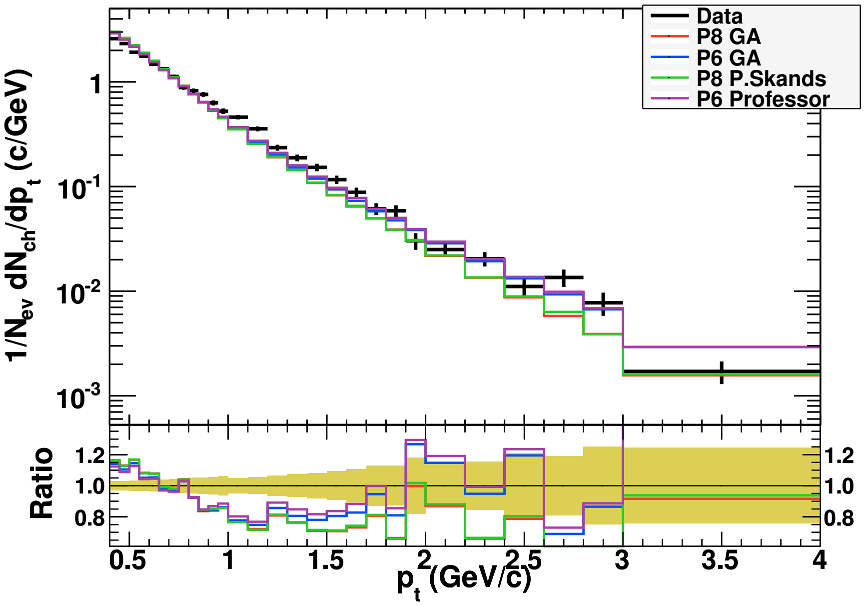


Figure 5.20.: Charged particle transverse momentum distributions from CDF at $\sqrt{s} = 630$ GeV in $|\eta| < 1$.

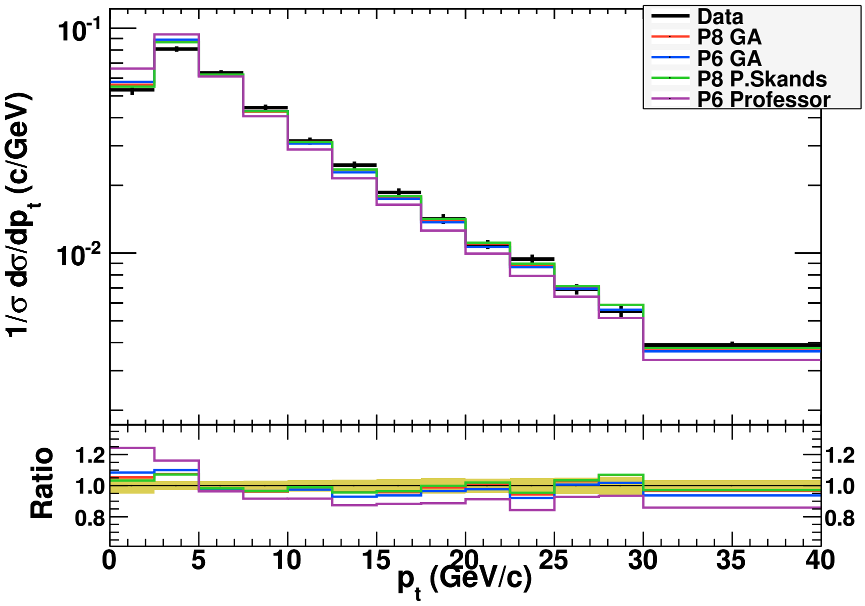


Figure 5.21.: Z_{p_T} distribution in Drell-Yan events from DØ at $\sqrt{s} = 1960$ GeV.

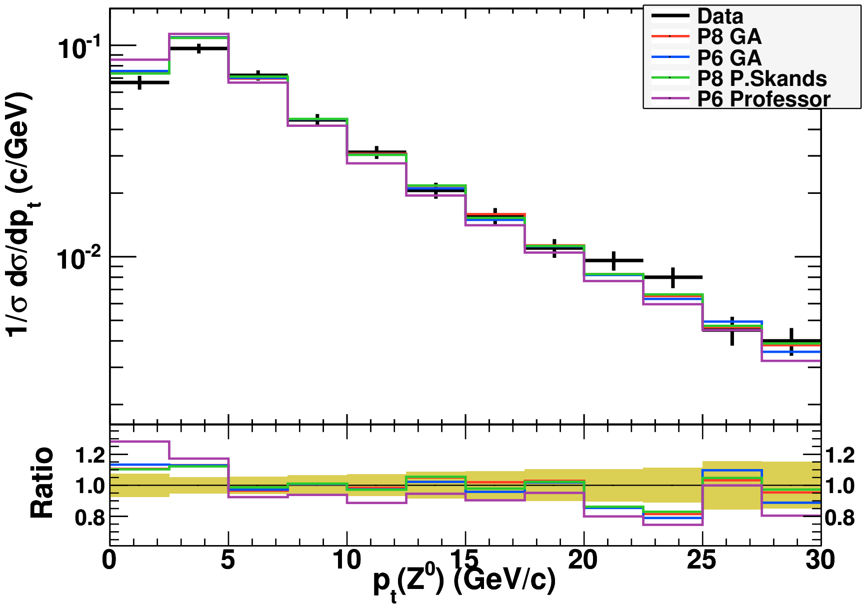


Figure 5.22.: Z_{p_T} distribution in Drell-Yan events at DØ at $\sqrt{s} = 1960$ GeV for events where the rapidity of Z is greater than 2.

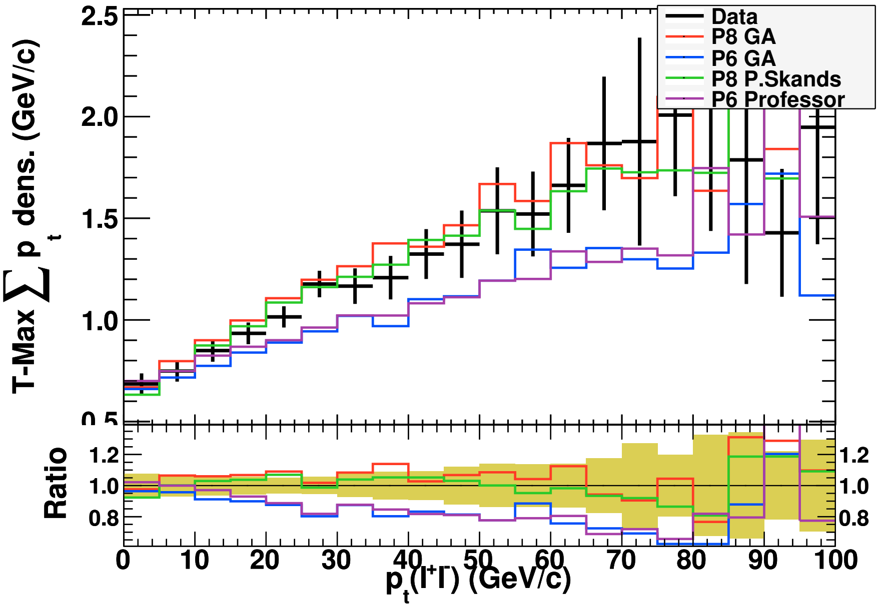


Figure 5.23.: Trans-Max region $\sum p_{\mathrm{T}}$ density in Drell-Yan events at CDF at $\sqrt{s} = 1960$ GeV.

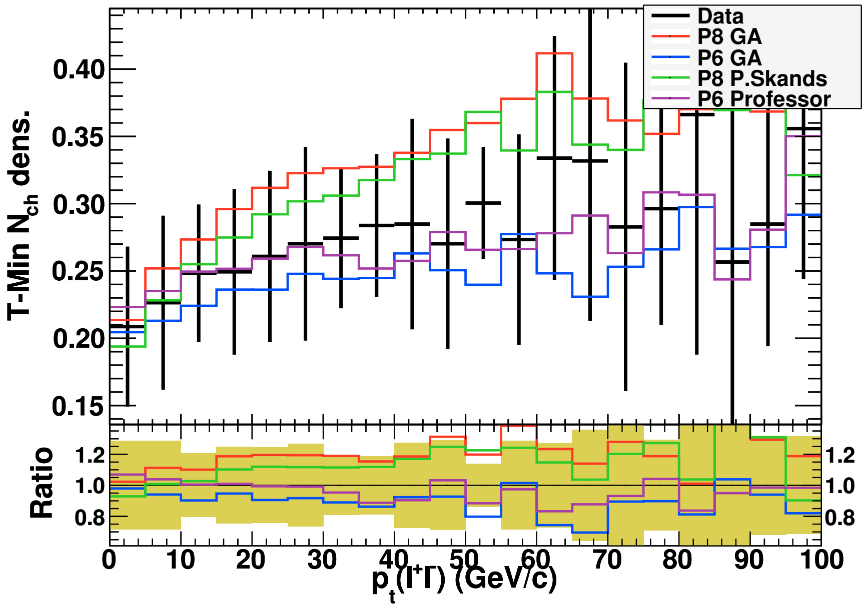


Figure 5.24.: Trans-Min region $\sum N_{\mathrm{ch}}$ density in Drell-Yan events at CDF at $\sqrt{s} = 1960$ GeV.

6. Data-MC Comparisons

A tune can change the data-MC agreement as much as permitted by the model used in the MC. However, different models have larger differences than tune-tune differences and different distributions. Thus comparing only tunes of a model is usually not enough and different tunes and models should be investigated. As shown in chapter 5 the tunes found by the GA improves the description of existing data and they are as good as the tunes found by the alternative methods. In order to investigate and compare the models, events generated by EPOS, PHOJET and PYTHIA were passed through the analysis library described in chapter 5. PYTHIA samples are generated using the tunes found by the GA for PYTHIA 8 using the new diffraction model and PYTHIA 6. All generators are set to generate diffractive, elastic and non-diffractive processes. Cross-sections of the different event types as calculated by PYTHIA, EPOS and PHOJET are listed in table 6.1. Figure 6.1 shows the evolution of the total cross-section with the center-of-mass energy. All generators describe the data well, but PHOJET overestimates a point at 1800 GeV. PHOJET predictions start to differ at about $\sqrt{s} = 500$ GeV. On the other hand EPOS and PYTHIA are within 10% of each other up to very high energies. Both PYTHIA versions are consistent with the cross-section predictions and have exactly the same values up to numerical precision.

6.1. Multiplicities

The $dN_{\text{ch}}/d\eta$ data from UA5 at 200 GeV as given in figure 6.2 are described within 10% at most places by EPOS, PHOJET and PYTHIA 6 and 8. The multiplicity distribution in the $|\eta| < 1.5$ region is given in figure 6.3 has differences at low multiplicity. Aside from EPOS, all other models start to deviate from the measurement below a multiplicity of 10, going beyond 20%.

Figure 6.4 shows the $dN_{\text{ch}}/d\eta$ distributions at 900 GeV. The PHOJET distribution is above the data, worse compared to 200 GeV. PYTHIA 8 has the best overall description. In the central region both EPOS and PYTHIA 6 have good descriptions, EPOS slightly overestimating in the high-η range. Figure 6.5 shows the same distribution with contributions from different event types as defined by the generators for PYTHIA 6 and 8. Although both have the same single and double-diffractive cross-sections, due to the new diffractive model in PYTHIA 8, the diffractive contributions are less than they are for PYTHIA 6. Figure 6.6 shows the same distribution for EPOS and PHOJET. The non-diffractive-only component of PHOJET has a very good agreement with the data, however with the addition of diffractive components it overshoots the data. This could be an indication of a problem in tuning of PHOJET 1.12. An interesting observation is

Energy	**PHOJET** (mbarn)				**EPOS** (mbarn)			
	σ_{tot}	σ_{dd}	σ_{sd}	σ_{nd}	σ_{tot}	σ_{dd}	σ_{sd}	σ_{nd}
1960 GeV	83.09	3.69	10.59	48.30	76.16	12.42	9.43	34.54
1800 GeV	81.50	3.68	10.64	47.32	74.92	12.32	9.34	34.01
900 GeV	69.20	3.53	10.53	39.98	65.62	11.49	8.72	29.97
630 GeV	63.63	3.42	10.33	36.72	61.33	11.12	8.44	28.04
546 GeV	61.60	3.37	10.25	35.51	59.83	10.97	8.32	27.34
200 GeV	50.75	2.71	8.92	29.99	50.36	10.00	7.58	22.85

Energy	**PYTHIA 6** (mbarn)				**PYTHIA 8** (mbarn)			
	σ_{tot}	σ_{dd}	σ_{sd}	σ_{nd}	σ_{tot}	σ_{dd}	σ_{sd}	σ_{nd}
1960 GeV	73.98	7.45	12.472	39.03	73.98	7.451	12.472	39.03
1800 GeV	72.98	7.34	12.388	38.48	72.97	7.334	12.388	38.49
900 GeV	65.35	6.40	11.677	34.44	65.36	6.399	11.678	34.45
630 GeV	61.78	5.93	11.298	32.61	61.79	5.931	11.298	32.62
546 GeV	60.42	5.75	11.143	31.93	60.42	5.746	11.142	31.93
200 GeV	51.90	4.49	10.002	27.85	51.90	4.491	10.002	27.86

Table 6.1.: Cross-sections calculated by PHOJET, EPOS, PYTHIA 6 and 8 at different energies.

that unlike other generators, EPOS predicts more double-diffractive events than single-diffractive events. A possible explanation could be that EPOS does not have a clear definition of the event type as other generators. The event type is deduced by looking at the internal particle numbering schemes which could result in differences with respect to other generators in the definition of the event type.

As shown in figure 6.7, the discrepancy at 200 GeV in the low multiplicity region of the charged particle multiplicity distributions in $|\eta| < 1.5$ becomes more significant at 900 GeV. PYTHIA 8 overestimates the peak by about 20% and PHOJET overshoots at the tails. The description of the low multiplicity range is still a problem for all generators, though PYTHIA 6 describes data at 900 GeV better than it does at 200 GeV.

Figure 6.8 shows the contributions from different event types as defined by the generator for PYTHIA 6 and 8 within $|\eta| < 1.5$ at 900 GeV. In events generated by PYTHIA 6, contributions from single and double-diffractive events are concentrated on the multiplicities $N_{\text{ch}} < 10$ and compensate the non-diffractive part. For PYTHIA 8, double-diffractive events have a very small contribution and the single-diffractive contribution is smaller and more spread than for PYTHIA 6. Moreover the non-diffractive part is about 25% higher than for PYTHIA 6 at the peak and slightly shifted towards higher multiplicity. Due to the difference between the diffractive events in PYTHIA 8, the rapid fall in the non-diffractive peak is not compensated as it is done for PYTHIA 6. At higher-η the description of the tails by PYTHIA 8 improves. Appendix B shows further plots.

Figure 6.9 shows the contributions from different event types for PHOJET and EPOS.

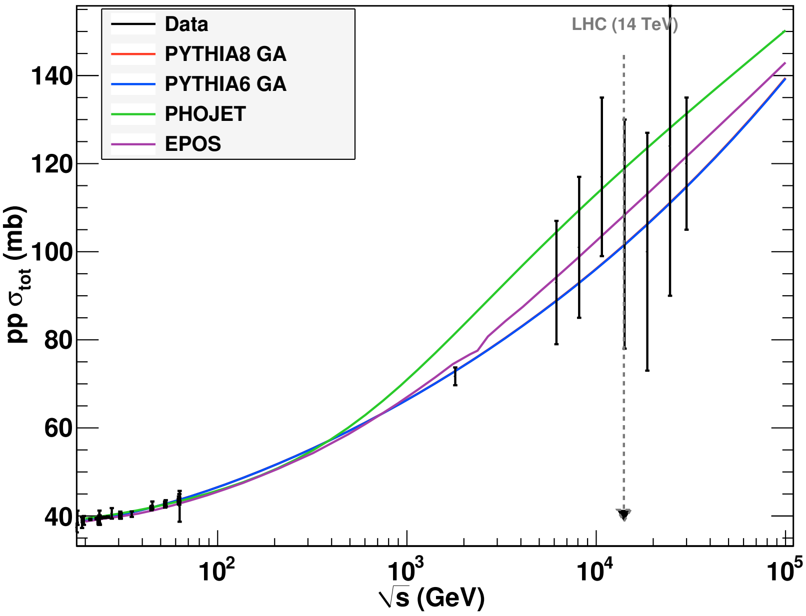


Figure 6.1.: Evolution of total *pp* cross-section value with the center-of-mass energy as calculated by EPOS, PHOJET, PYTHIA 6 and 8. PYTHIA 8 and 6 curves lie perfectly on top of each other. Data points are taken from [18].

In events generated by PHOJET, single and double-diffractive events have longer tails than for other generators, however their contributions are not enough to describe the data in first three bins. EPOS has the largest diffractive contributions and has a good description of the low multiplicity range.

EPOS has overall good agreement with the data, PHOJET overestimates high multiplicity events and $dN_{\mathrm{ch}}/d\eta$, PYTHIA 6 and 8 have a very good agreement with $dN_{\mathrm{ch}}/d\eta$ distributions and acceptable agreement with the multiplicity data. Appendix B contains comparisons with UA5 data at various η ranges.

Recently the ATLAS collaboration has measured $dN_{\mathrm{ch}}/d\eta$ distribution of particles with $p_{\mathrm{T}} > 0.5$ GeV for the events with at least one particle with $p_{\mathrm{T}} > 0.5$ in $|\eta| < 2.5$ at 900 GeV using approximately $\mu\mathrm{b}^{-1}$ data [74]. Figure 6.10 shows the data corrected for the trigger and detector effects together with predictions from generators. The PYTHIA 6 GA tune underestimates the data by about 15% in all η ranges. PYTHIA 8 and PHOJET both remain about 5% below the data and EPOS is within the error bars, slightly underestimating data in the central region.

CMS collaboration published $dN_{\mathrm{ch}}/d\eta$ measurements at $\sqrt{s} = 900$ GeV and 2360 GeV corrected to non-single-diffractive events [75]. Figures 6.11 and 6.12 show the results at $\sqrt{s} = 900$ GeV and $\sqrt{s} = 2360$ GeV, respectively. At $\sqrt{s} = 900$ GeV, PHOJET and EPOS overestimate and PYTHIA 8 underestimates the data. PYTHIA 6 is in agreement with data. At $\sqrt{s} = 2360$ GeV EPOS and PHOJET stay within the errors and PYTHIA 6 and 8 underestimates the data.

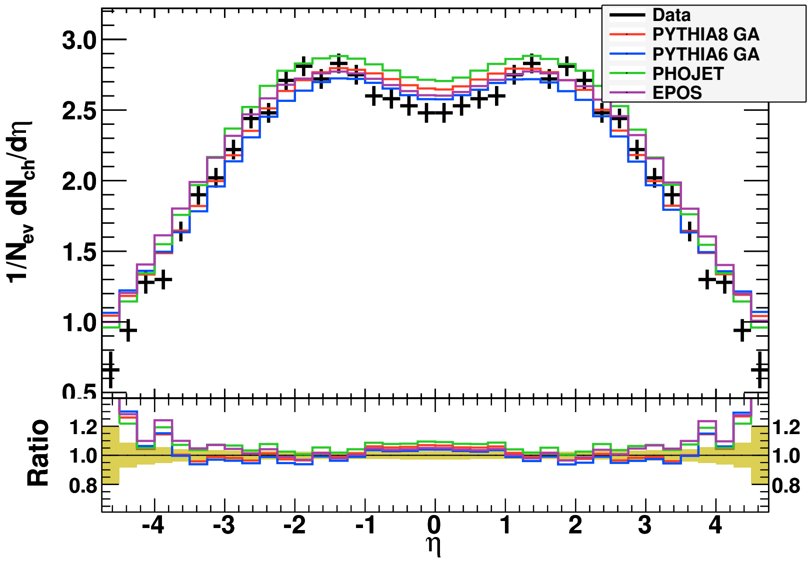


Figure 6.2.: Charged particle distributions at 200 GeV from UA5. Yellow areas represent the error on data.

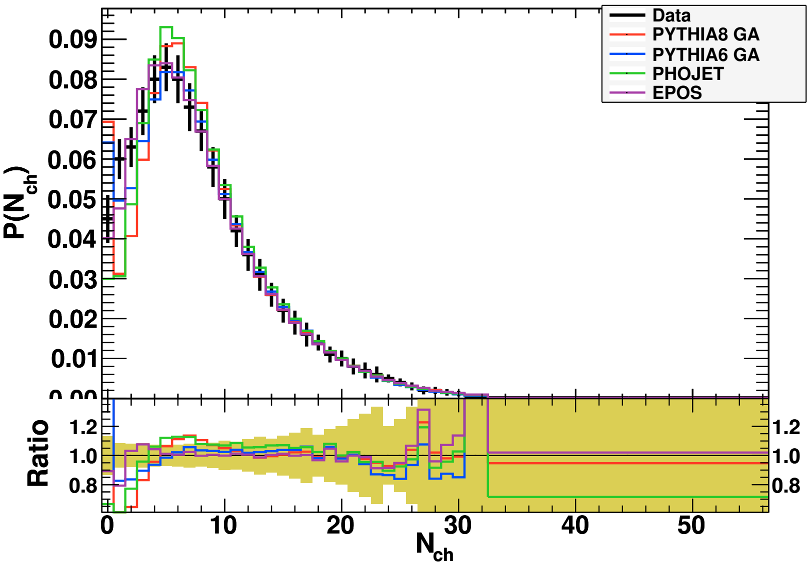


Figure 6.3.: Charged particle multiplicity distribution for $|\eta| < 1.5$ at $\sqrt{s} = 200$ GeV from UA5. Yellow areas represent the error on data.

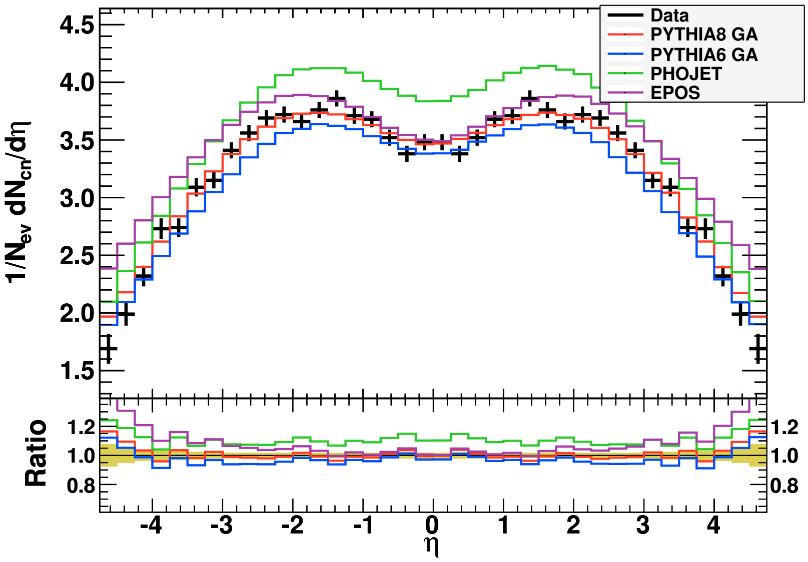


Figure 6.4.: Charged particle pseudorapidity distributions at 900 GeV from UA5. Yellow areas represent the error on data.

The evolution of the central charged particle multiplicities in non-single-diffractive events, as defined by generators, are given in figure 6.13. UA5 and UA1 data lie slightly below all MC curves. PHOJET overestimates the data. EPOS, PYTHIA 8 and 6 stays within the measurements of STAR, CDF, ALICE and CMS collaborations. CMS data points are corrected for the single-diffractive contributions and non-single-diffractive inefficiencies using PYTHIA 6 D6T tune [75], other data points are selected using a two-arm trigger and not corrected for event types. Figure 6.14 shows the same distribution for events selected by requiring at least one charged particle in $1.5 < |\eta| < 5.5$ in both η hemispheres. A similar event selection is used in both UA1 and UA5 experiments [67, 78]. It can be clearly seen from both plots that curves differ depending on their models. PHOJET curve has the least change between trigger selected and event type selected plots and the PYTHIA 8 curve, on the other hand, has the largest difference. Overall agreement with the data is increased significantly in two-arm trigger selected events. Figure 6.15 shows the same plot with several PYTHIA tunes for events with at least one particle with $E > 3$ GeV in $2.9 < |\eta| < 5.2$ for both η hemispheres and a charged particle with $p_{\mathrm{T}} > 0.1$ GeV in $|\eta| < 2.5$, similar to the CMS event selection. All curves except PHOJET agree with the data up to 900 GeV with differences with respect to each other in the order of 5%. The same plot for non-single-diffractive events is shown in figure 6.16. Plots show varying differences. PHOJET and the PYTHIA 6 D6T tune have relatively smaller changes with respect to other tunes and models between figures 6.15 and 6.16. Since the CMS collaboration used PHOJET and PYTHIA 6 D6T tunes to estimate the corrections to data, there can be some underestimated systematics which become significant at higher energies. In order to avoid dependencies to tunes

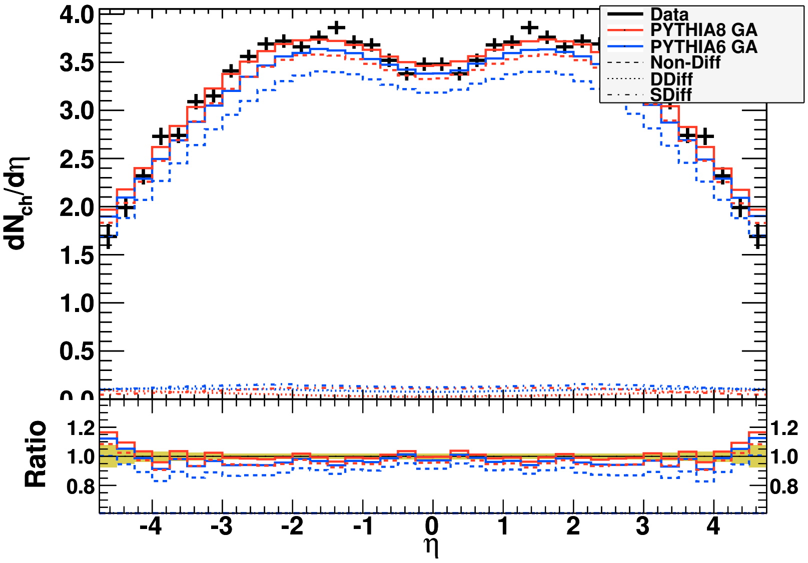


Figure 6.5.: Charged particle pseudorapidity distribution with contributions from event types for PYTHIA 6 and 8. Red and blue lines represent PYTHIA 6 and 8, respectively. Dashed lines show non-diffractive, dots show double-diffractive and dash-dots show single-diffractive contributions. Data are from UA5 at $\sqrt{s} = 900$ GeV. Yellow areas represent the error on data.

or models experiments should also publish the data corrected for detector and trigger acceptances and efficiencies with as small model dependence as possible. In that way, different models can be studied with a lesser possible bias and more reliable information can be deduced from the data.

6.2. Transverse Momentum

The average p_{T} as a function of the charged particle multiplicity in the event at 1960 GeV from CDF is shown in figure 6.17. PHOJET has a less steep distribution quickly diverging from the measurements as N_{ch} increases. This behavior might be related to the rather old parton distribution functions (GRV98) used in PHOJET. PYTHIA 6 and 8 have a good description of the shape and usually stay within 5% of the data. At low multiplicity the difference between data increases up to 10%. EPOS has a different shape but stays within 5% of the data. Figure 6.18 shows the the same average p_{T} distribution for $p_{\mathrm{T}} > 0$. Variations between the generators reach about 20% at low multiplicities. Differences between the models without the p_{T} cut is an indication that extrapolating the data to outside of the measured phase space is dangerous and it should be avoided if possible.

The average transverse momentum distribution at $\sqrt{s} = 900$ GeV for charged particles with $p_{\mathrm{T}} > 0.5$ and $|\eta| < 2.5$ as measured by the ATLAS detector and corrected for

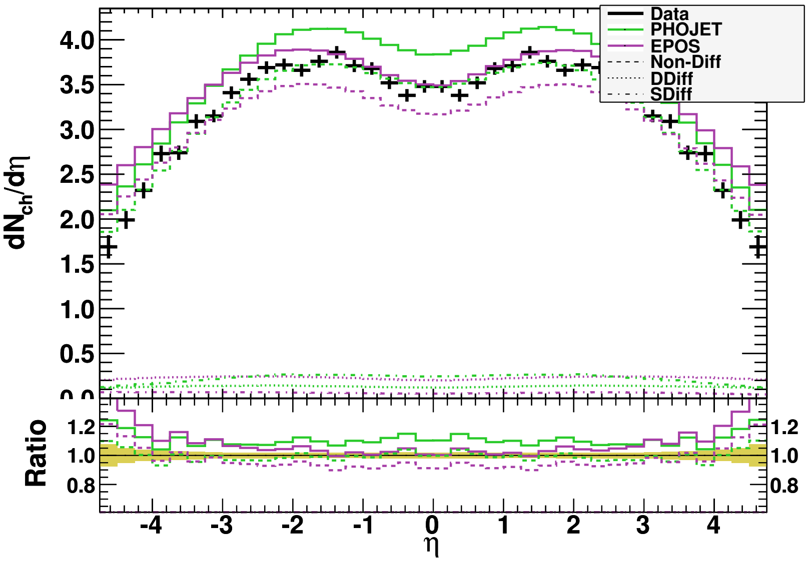


Figure 6.6.: Charged particle pseudorapidity distribution with contributions from event types for EPOS and PHOJET. Green and violet curves represent PHOJET and EPOS, respectively. Dashed lines show non-diffractive , dots show double-diffractive and dash-dots show single-diffractive contributions. Data are from UA5 at $\sqrt{s} = 900$ GeV. Yellow areas represent the error on data.

detector and trigger effects is given in figure 6.19. None of the generators describe the distribution correctly. PYTHIA 6 underestimates the data but has a shape similar to it. EPOS overestimates the data for multiplicities less than 10 and underestimates above. Both PYTHIA 6 and EPOS stay within 5% of the data. PHOJET underestimates the data for multiplicities greater than 5 particles reaching about 10% difference at high multiplicities. PYTHIA 8 is as much as 8% below data for $N_{\mathrm{ch}} < 30$.

Transverse momentum distributions at 630 GeV and 1800 GeV are given in figures 6.20 and 6.21, respectively. PHOJET, PYTHIA 6 and 8 usually stay within 20% of the measurement. However EPOS has a better agreement and stays within 10% of the measurement. As can be seen from figure 6.22, at 1960 GeV EPOS has better than 2% agreement for $p_{\mathrm{T}} < 2.5$ GeV. PHOJET and PYTHIA 8 are about 10% below the data, PYTHIA 6 overshoots the data at higher p_{T}.

The evolution of the average p_{T} at different center-of-mass energies is given in figure 6.23. All generators predict higher average p_{T} than the UA1 and ISR measurements. EPOS shows the largest deviation from the UA1 data. PYTHIA6 significantly diverges from the data at low $\sqrt{s}$ values. However, it describes CDF and CMS data well. PHOJET produces the closest values to ISR and UA1 data but it is below the recent CMS measurement at 2360 GeV. PYTHIA 8 has a good description of CDF and CMS data and close values to UA1 and ISR measurements.

In general all generators have similar descriptions of the p_{T} distributions. However

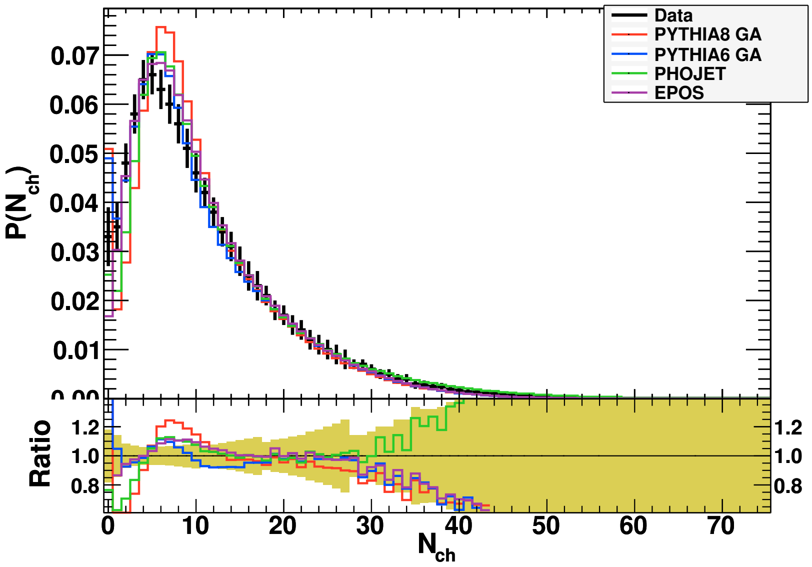


Figure 6.7.: Charged particle multiplicity distribution at 900 GeV in $|\eta| < 1.5$ from UA5. Yellow areas represent the error on data.

EPOS has the best description of the low momentum part of the distributions.

Although GA tunes have good descriptions of multiplicity data and acceptable descriptions of transverse momentum at $\sqrt{s} = 900$ GeV, the new ATLAS data are not well described. This could be an indication of a not well constrained parameter space and of the need for more data covering different parts of the phasespace.

The new diffractive model in PYTHIA 8 improves the definition of both the momentum and multiplicity distributions and it has a good description of data. It is slightly worse than PYTHIA 6 at low multiplicies. The new model in the EPOS has the overall best description of the existing data. Among all generators, PHOJET has the worst description of the data. This could be attributed to old PDF sets. Updating it to use recent PDF sets may improve its performance. Also there were some indications of possible bugs in the PHOJET code. It is not possible to favor one model over the others with the existing data.

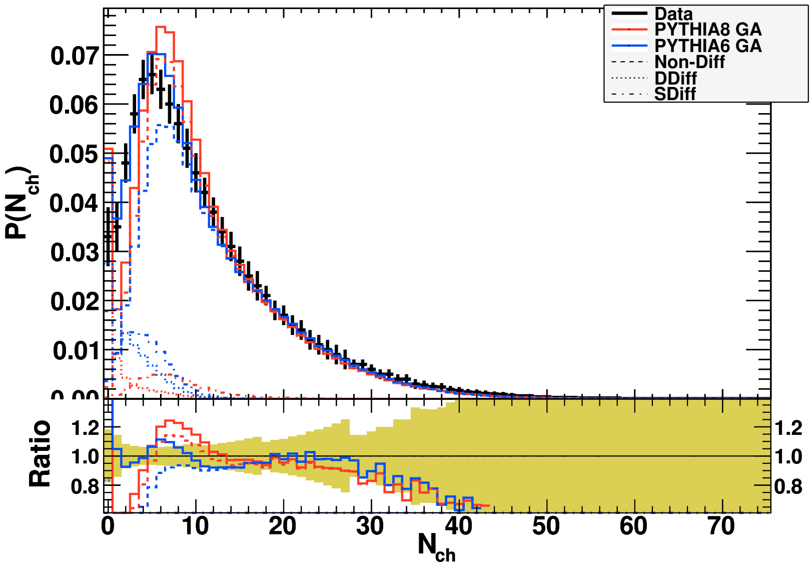


Figure 6.8.: Charged particle multiplicity distribution with contributions from different event types for PYTHIA 6 and 8 at 900 GeV in $|\eta| < 1.5$. Data are from UA5 at $\sqrt{s} = 900$ GeV. Yellow areas represent the error on data.

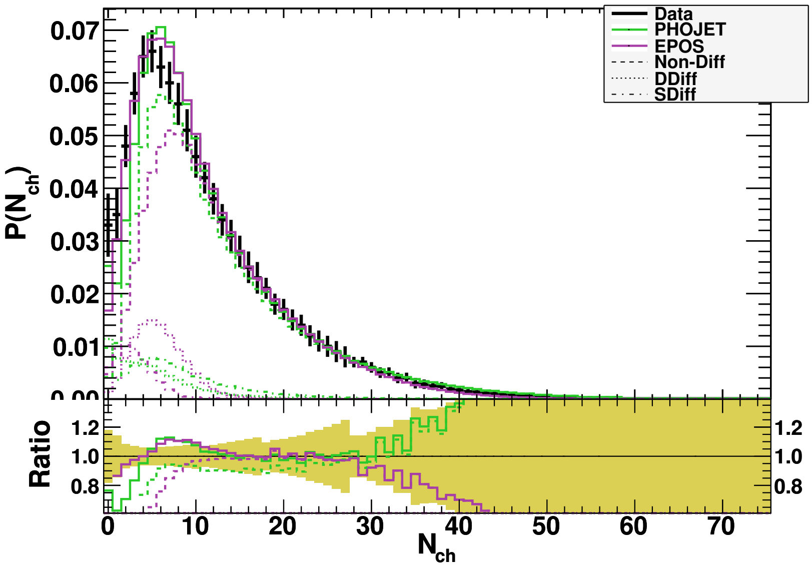


Figure 6.9.: Charged particle multiplicity distribution with contributions from different event types for PHOJET and EPOS at 900 GeV within $|\eta| < 1.5$. Data are from UA5 at $\sqrt{s} = 900$ GeV. Yellow areas represent the error on data.

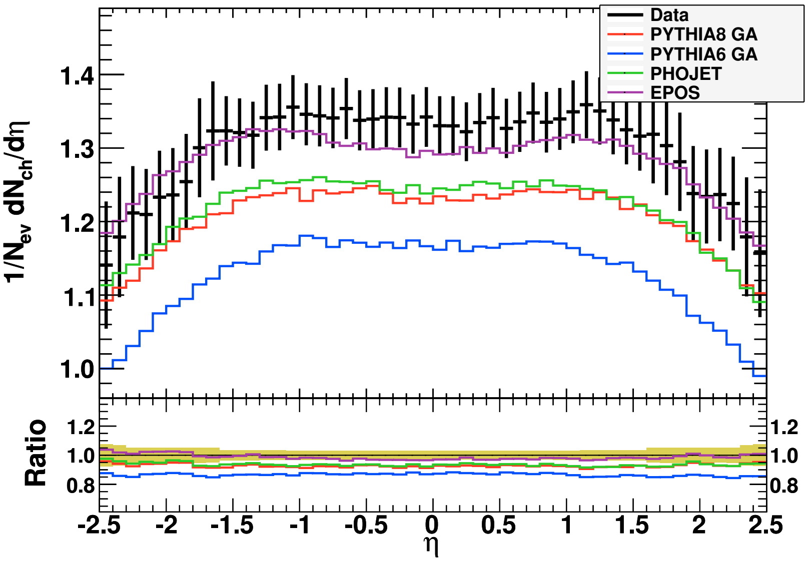


Figure 6.10.: ATLAS detector pseudorapidity measurement of charged particles with $p_{\mathrm{T}} > 0.5$ GeV and $|\eta| < 2.5$ at $\sqrt{s} = 900$ GeV. The data are corrected for detector and trigger effects [74].

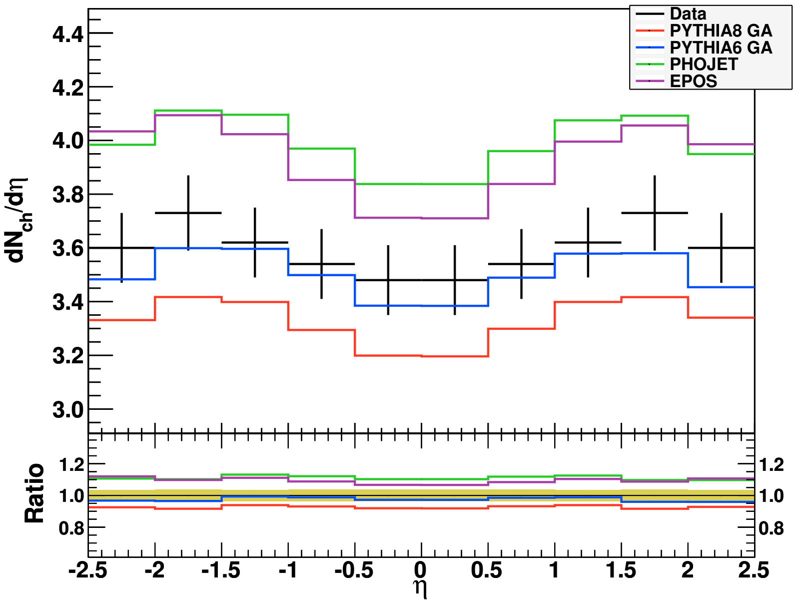


Figure 6.11.: CMS collaboration charged particle pseudorapitidty measurement at $\sqrt{s} = 900$ GeV. Data is corrected to NSD events using PYTHIA 6 D6T tune [75].

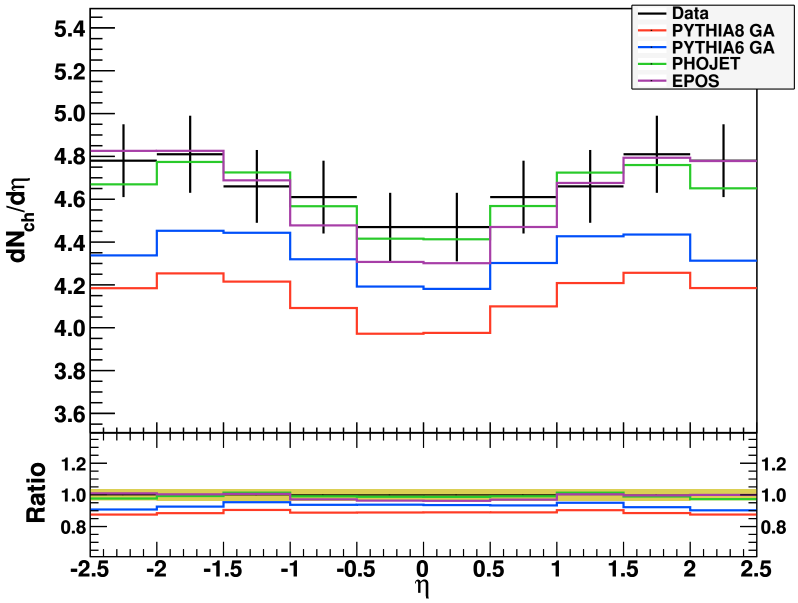


Figure 6.12.: CMS collaboration charged particle pseudorapitidty measurement at $\sqrt{s} = 2360$ GeV. Data is corrected to NSD events using PYTHIA 6 D6T tune [75].

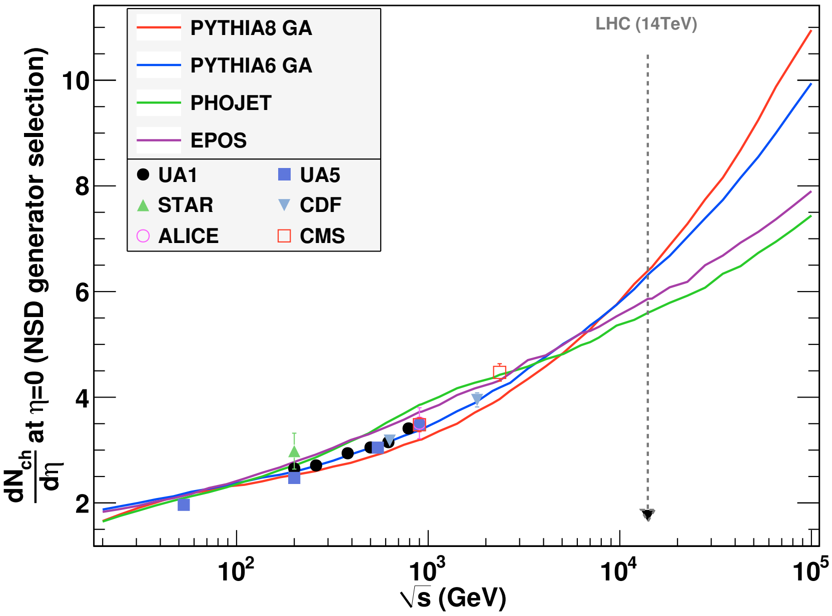


Figure 6.13.: Evolution of central charged particle multiplicity of non-single-diffractive events with $\sqrt{s}$. Data points are from [60, 67, 75–78].

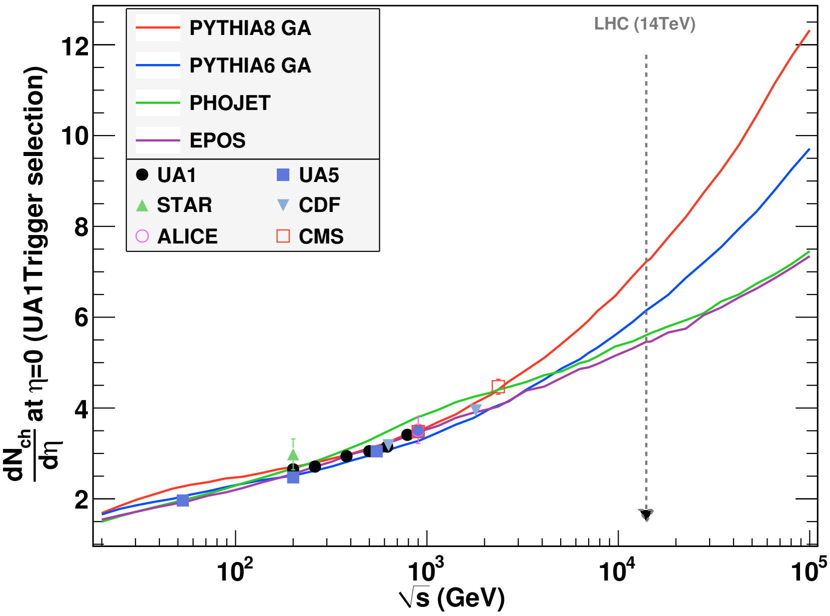


Figure 6.14.: Central charged particle multiplicity for events with at least one charged particle in $1.5 < |\eta| < 5.5$ in both η hemispheres at different $\sqrt{s}$. Data points are from [60, 67, 75–78].

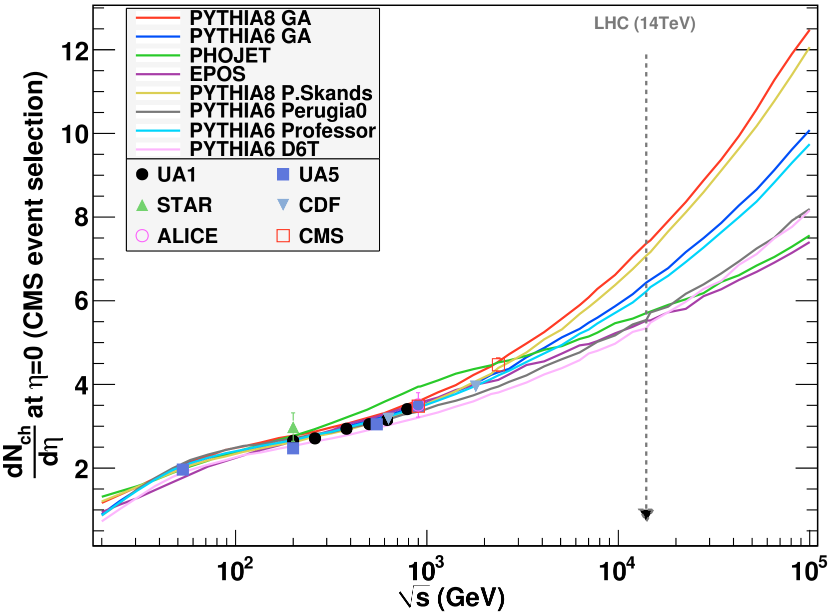


Figure 6.15.: Central charged particle multiplicity for events with at least one particle with $E > 3$ GeV in $2.9 < |\eta| < 5.2$ in both hemispheres and at least one charged particle with $p_T > 0.1$ GeV in $|\eta| < 2.5$ at different $\sqrt{s}$. Data points are from [60, 67, 75–78].

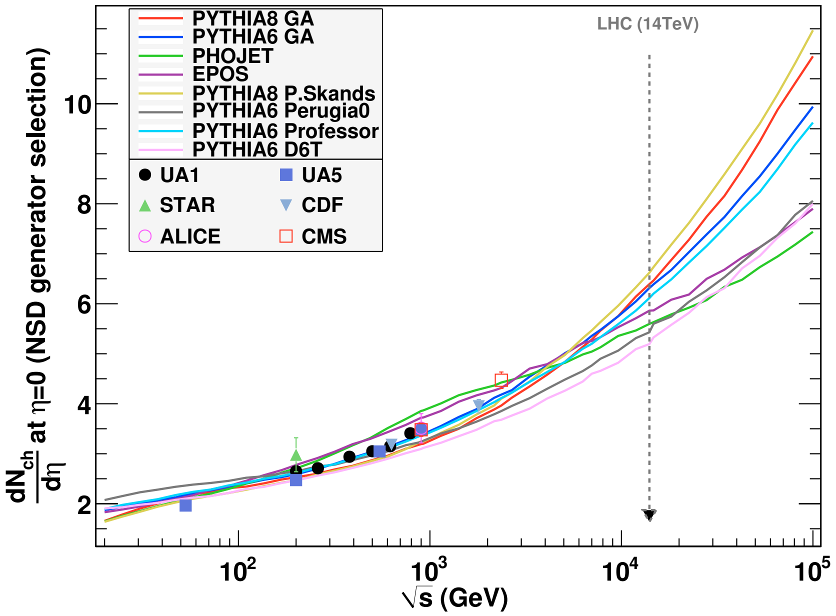


Figure 6.16.: Central charged particle multiplicity of non-single-diffractive events at different $\sqrt{s}$. Data points are from [60, 67, 75–78].

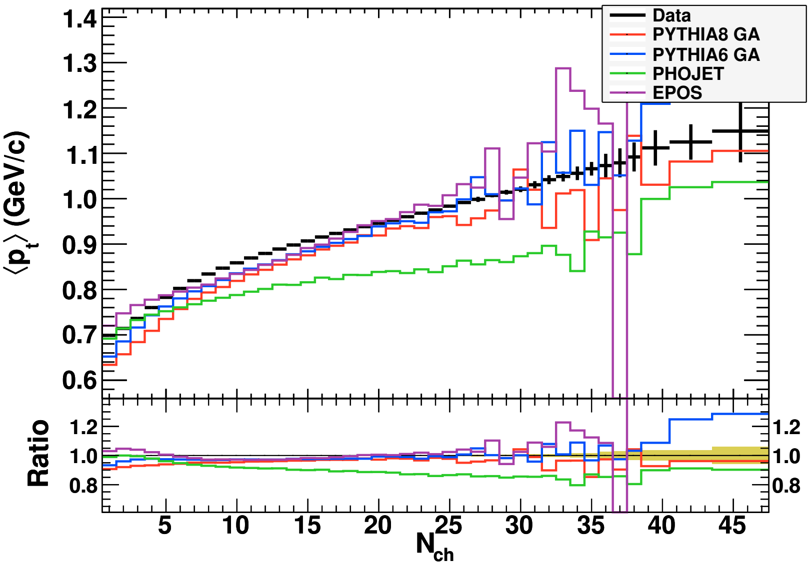


Figure 6.17.: Average p_T distributions for $|\eta| < 1$ and $p_T > 0.4$ at 1960 GeV from CDF. Yellow areas represent the error on data.

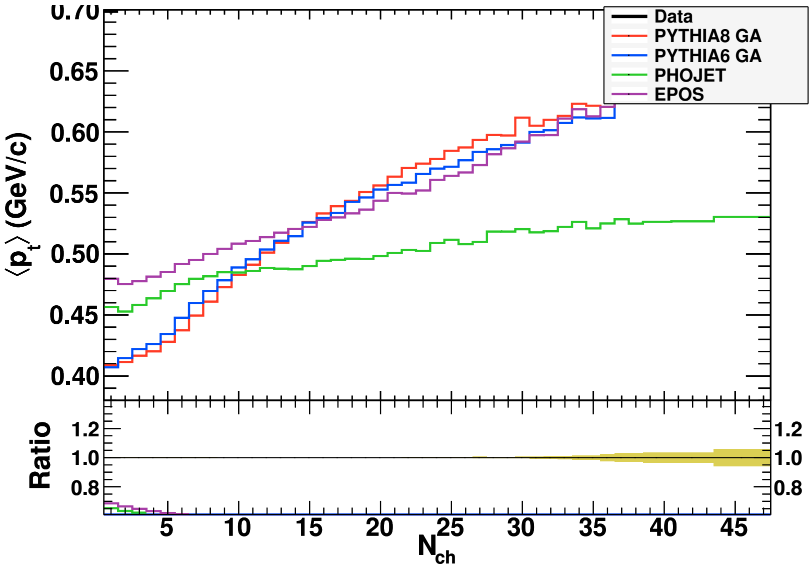


Figure 6.18.: Average p_T distributions in $|\eta| < 1$, $p_T > 0$ GeV at $\sqrt{s} = 1960$ GeV.

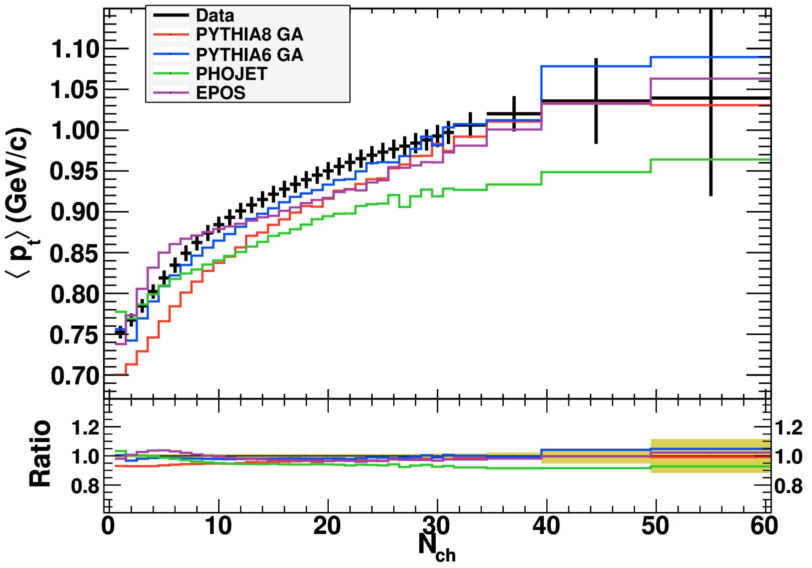


Figure 6.19.: Average p_T at different event multiplicities for particles with $p_T > 0.5$ and $|\eta| < 2.5$ at $\sqrt{s} = 900$ GeV as measured by ATLAS collaboration [74].

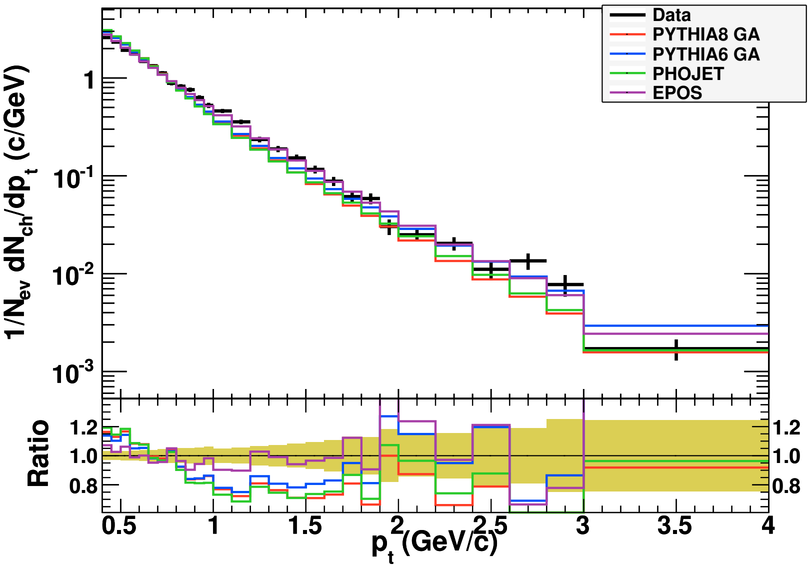


Figure 6.20.: Transverse momentum distributions in $|\eta| < 1.$ and $p_{\mathrm{T}} > 0.4$ GeV at $\sqrt{s} = 630$ GeV from CDF. Yellow areas represent the error on data.

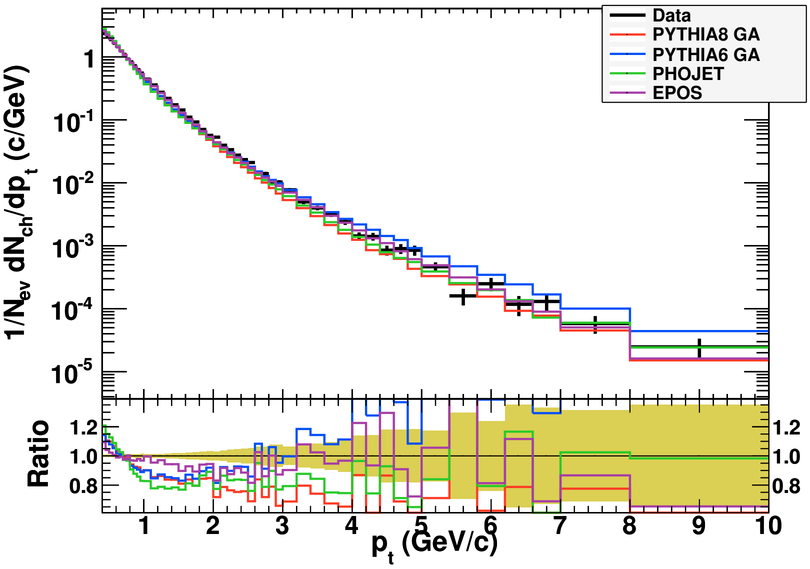


Figure 6.21.: Transverse momentum distributions in $|\eta| < 1.$ and $p_{\mathrm{T}} > 0.4$ GeV at $\sqrt{s} = 1800$ GeV from CDF. Yellow areas represent the error on data.

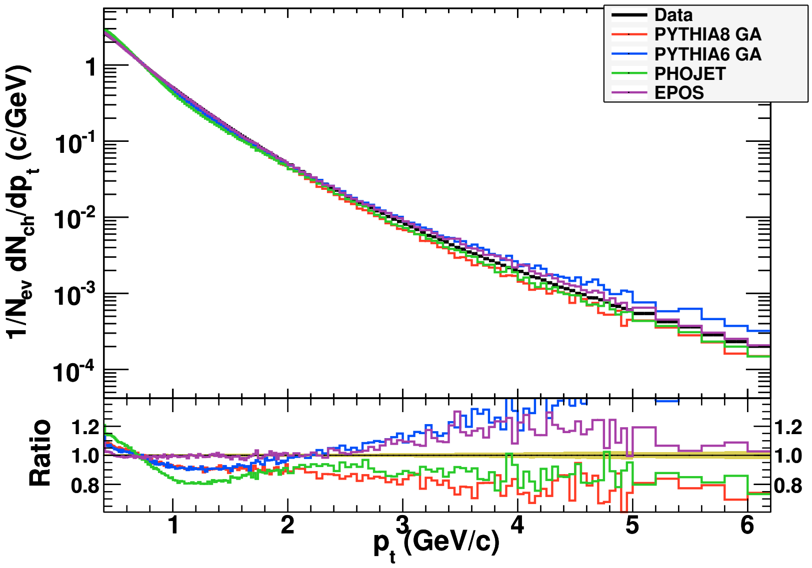


Figure 6.22.: Transverse momentum distributions for $|\eta| < 1.$ and $p_{\mathrm{T}} > 0.4$ GeV at $\sqrt{s} = 1960$ GeV from CDF. Yellow areas represent the error on data.

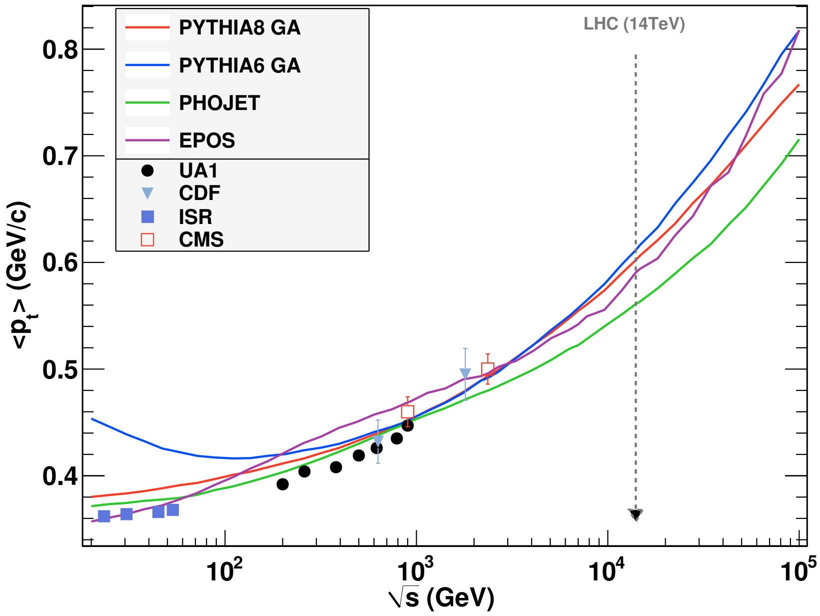


Figure 6.23.: Evolution of average charged particle p_{T} as a function of center-of-mass energy. Data points are taken from [57, 75, 78, 79].

7. Predictions for LHC energies

In September 2008, a faulty connection between magnets of the LHC caused a malfunction and operation of the LHC was delayed until November 2009. The LHC experiments collected some data at $\sqrt{s} = 900$ GeV and 2360 GeV slightly surpassing Tevatron energies before the 2009 annual shutdown of CERN. The ATLAS experiment collected about $\sim 9\mu\mathrm{b}^{-1}$ at 900 GeV. Almost all MC generators can describe existing data up to some extend, but since generators predict different distributions outside the measured phase space even measurements at 900 GeV provide important information. Data above 2 TeV will provide information never measured before.

In this chapter basic observables at generator level for different energies, using basic detector acceptance and triggering requirements of the ATLAS detector will be presented. In addition some distributions using full detector and trigger simulation at 900, 2360 and 7000 GeV will be shown at the end. The basic triggering requirement is at least one charged particle in the Minimum Bias Trigger Scintillators (MBTS) range, covering $2.12 < |\eta| < 3.85$, which would correspond to a L1 MBTS trigger item. Level 1 MBTS trigger items are labeled depending on the number of hits on the planes. If one of the planes has a hit the trigger is called *L1_MBTS_1*, if there are at least one hit on both planes it is named *L1_MBTS_1_1* and if there are at least two hits on one of the planes then it is named *L1_MBTS_2*. For generator level plots, event selection requires L1_MBTS_1, at least 3 particles with $p_\mathrm{T} > 0.2$ GeV in $|\eta| < 2.5$ and at least one charged particle with $p_\mathrm{T} > 0.5$ GeV and $|\eta| < 2.5$. Unless specified only particles with $p_\mathrm{T} > 0.5$ GeV and $c\tau > 10\,\mathrm{mm}$ are used in the generator level plots.

7.1. Multiplicity distributions

One of the very first measurements using the data will be $dN_\mathrm{ch}/d\eta$ distributions. Figure 7.1 shows expected $dN_\mathrm{ch}/d\eta$ distributions from the ATLAS detector at 900 GeV for events triggered by MBTS_1 and which have a track with $p_\mathrm{T} > 0.5$ GeV in $|\eta| < 2.5$. Both PYTHIA versions have about the same shape and value, however EPOS and PHOJET predict about 5 to 10% more particles. Figure 7.2 shows the distributions at 2360 GeV. All generators are within 5% of each other with different shapes, PYTHIA predicting more charged particles than EPOS and PHOJET. In addition, PYTHIA 8 and PYTHIA 6 show slight differences. The difference between generators reaches 20% at 7 TeV and more than 25% at 14 TeV, as can be seen in figures 7.3 and 7.4. It can be clearly seen from the figures 7.1 to 7.4 that generators have different $\sqrt{s}$ extrapolations of the distributions.

The charged particle multiplicity distributions in $|\eta| < 2.5$ for particles with $p_\mathrm{T} > 0.5$ GeV at different energies are shown in figures 7.5 to 7.8. They show large variations

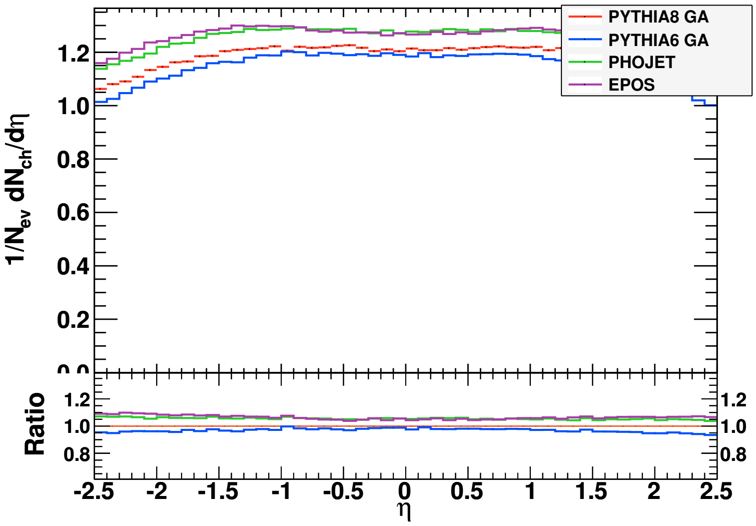


Figure 7.1.: Charged particle pseudorapidity distributions of particles with $p_{\mathrm{T}} > 0.5$ GeV at $\sqrt{s} = 900$ GeV for the events where the MBTS_1 trigger condition and the existence of at least one particle with $p_{\mathrm{T}} > 0.5$ inside $|\eta| < 2.5$ is satisfied. The bottom plot shows the ratio with respect to the PYTHIA 8 distribution.

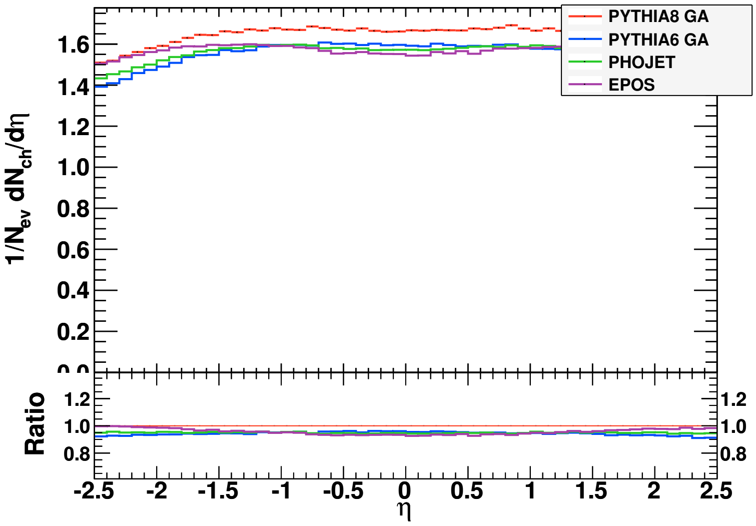


Figure 7.2.: Charged particle pseudorapidity distributions of particles with $p_{\mathrm{T}} > 0.5$ GeV at $\sqrt{s} = 2360$ GeV for the events where the MBTS_1 trigger condition and the existence of at least one particle with $p_{\mathrm{T}} > 0.5$ inside $|\eta| < 2.5$ is satisfied. The bottom plot shows the ratio with respect to the PYTHIA 8 distribution.

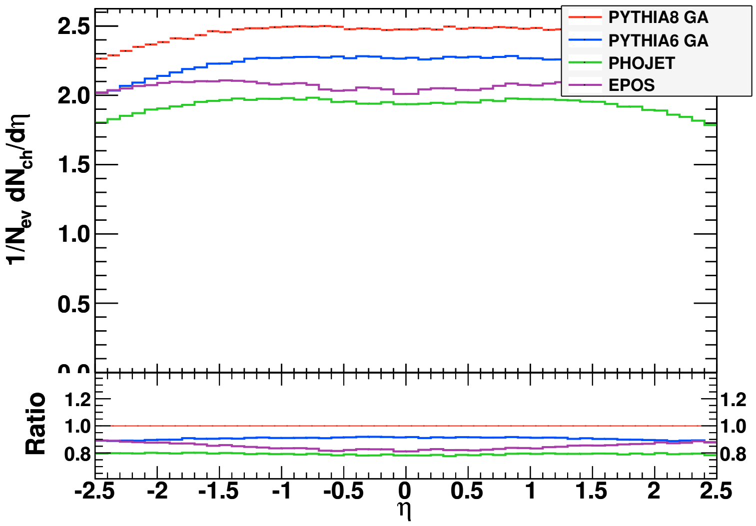


Figure 7.3.: Charged particle pseudorapidity distributions of particles with $p_T > 0.5$ GeV at $\sqrt{s} = 7$ TeV for the events where the MBTS_1 trigger condition and the existence of at least one particle with $p_T > 0.5$ inside $|\eta| < 2.5$ is satisfied. The bottom plot shows the ratio with respect to the PYTHIA 8 distribution.

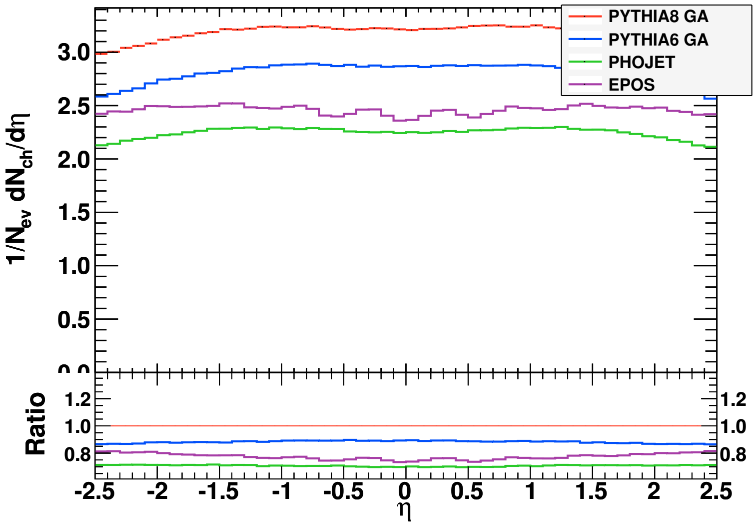


Figure 7.4.: Charged particle pseudorapidity distributions of particles with $p_T > 0.5$ GeV at $\sqrt{s} = 14$ TeV for the events where the MBTS_1 trigger condition and the existence of at least one particle with $p_T > 0.5$ inside $|\eta| < 2.5$ is satisfied. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

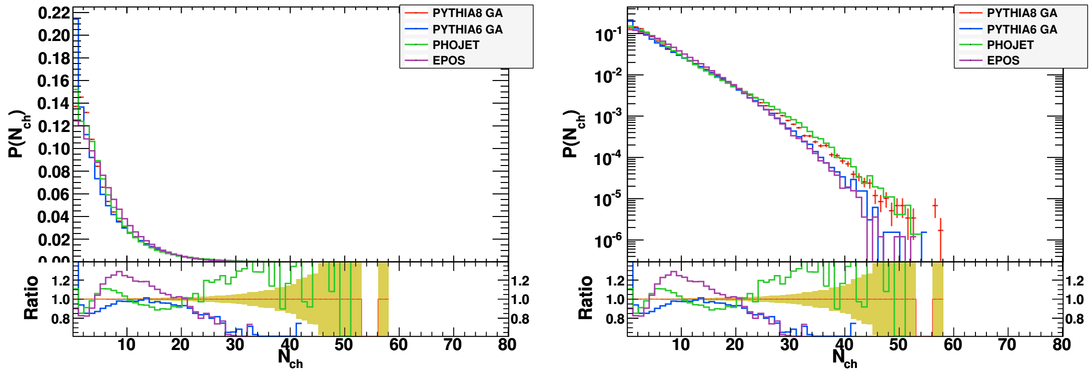


Figure 7.5.: Charged particle multiplicity for particles with $p_\mathrm{T} > 0.5$ GeV in $|\eta| < 2.5$ at $\sqrt{s} = 900$ GeV for events where the MBTS_1 triggered. The left plot is linear and the right plot is logarithmic scale. The bottom plots show the ratio with respect to the PYTHIA 8 distribution.

with respect to each other. Especially at high energies differences in the high multiplicity tails become large. Once the detector acceptances and efficiencies are moderately understood, the high multiplicity tails might provide some discrimination between the two extremes, EPOS and PYTHIA8, using 2360 GeV data.

7.2. Transverse Momentum Distributions

Figures 7.9 to 7.12 show the p_T distributions of the events satisfying the MBTS_1 using particles with $p_\mathrm{T} > 0.5$ GeV. Transverse momentum distributions depend on various factors in the underlying models of the generators, thus they are quite different from each other. The shape of the distribution and the extrapolation to higher center-of-mass energies differ quite significantly between the tunes and generators. Measurements of p_T distributions at different energies may provide a strong handle for model discrimination.

7.3. Average Transverse Momentum

Average $\langle p_\mathrm{T} \rangle$ measurement requires better understanding of the detector efficiencies and acceptances, thus it is usually produced after $dN_\mathrm{ch}/d\eta$, p_T and N_ch measurements. However $\langle p_\mathrm{T} \rangle$ is related to the many parameters in the models of the generators and is one of the important handles in tuning. Figures 7.13 to 7.16 show the average p_T versus charged particle multiplicity distributions for particles with $p_\mathrm{T} > 0.5$ GeV at different energies for events where the MBTS_1 triggered. For multiplicities greater than 10, generators usually stay within 5% of each other. PHOJET has a flatter behavior than the other generators, differing up to 15% at high multiplicities. The low multiplicity part of the distributions are different for all generators, resulting in differences of up to 20%. The extrapolation to higher energies are also slightly different. PYTHIA 8 has the

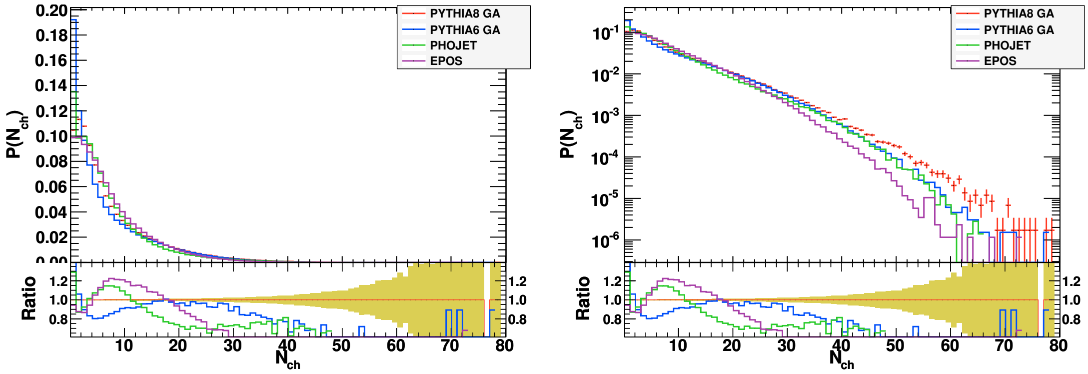


Figure 7.6.: Charged particle multiplicity for particles with $p_{\mathrm{T}} > 0.5$ GeV in $|\eta| < 2.5$ at $\sqrt{s} = 2360$ GeV for events where the MBTS_1 triggered. Linear scale is on the left, Logarithmic scale is on the right. The bottom plots show the ratio with respect to PYTHIA 8 distribution.

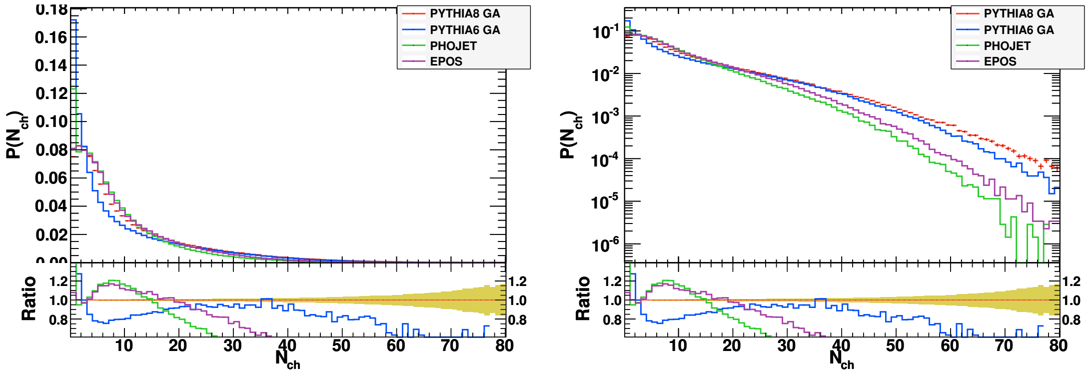


Figure 7.7.: Charged particle multiplicity for particles with $p_{\mathrm{T}} > 0.5$ GeV in $|\eta| < 2.5$ at $\sqrt{s} = 7$ TeV for events where the MBTS_1 triggered. Linear scale is on the left, Logarithmic scale is on the right. The bottom plots show the ratio with respect to PYTHIA 8 distribution.

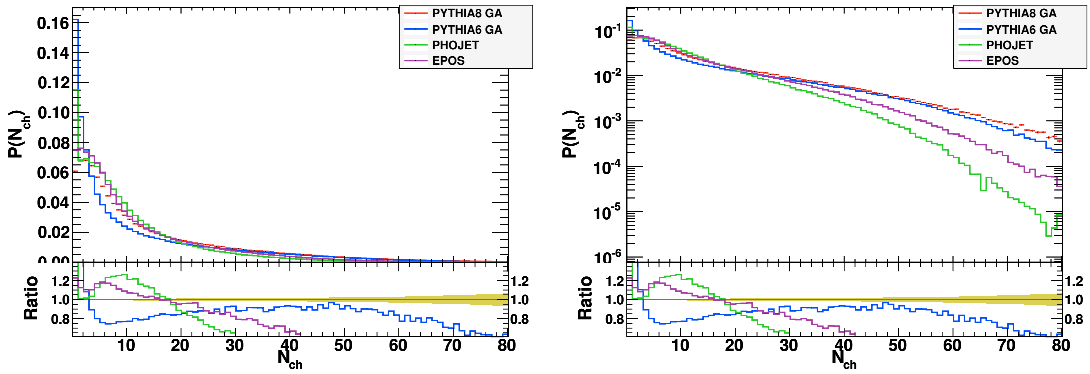


Figure 7.8.: Charged particle multiplicity of particles with $p_{\mathrm{T}} > 0.5$ GeV in $|\eta| < 2.5$ at $\sqrt{s} = 14$ TeV for events where the MBTS_1 triggered. Linear scale is on the left, Logarithmic scale is on the right. The bottom plots show the ratio with respect to PYTHIA 8 distribution.

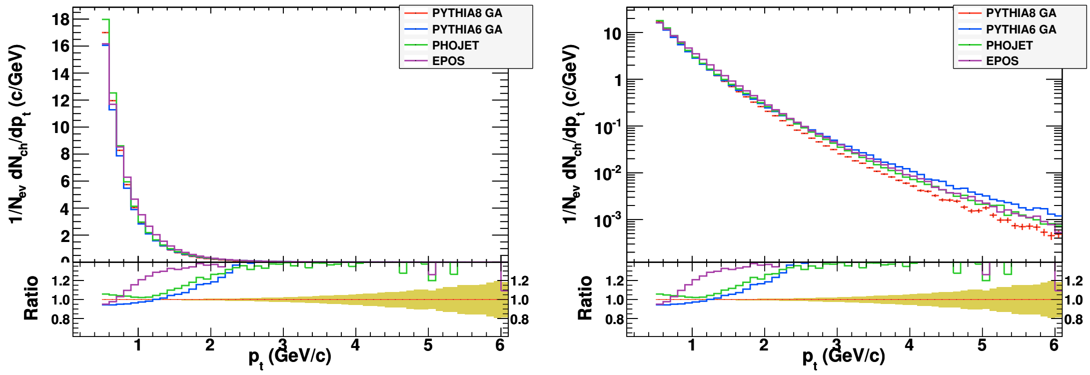


Figure 7.9.: Transverse momentum distribution for $|\eta| < 2.5$ at $\sqrt{s} = 900$ GeV for events where MBTS_1 triggered. Linear scale is on the left, Logarithmic scale is on the right. The bottom plots show the ratio with respect to PYTHIA 8 distribution.

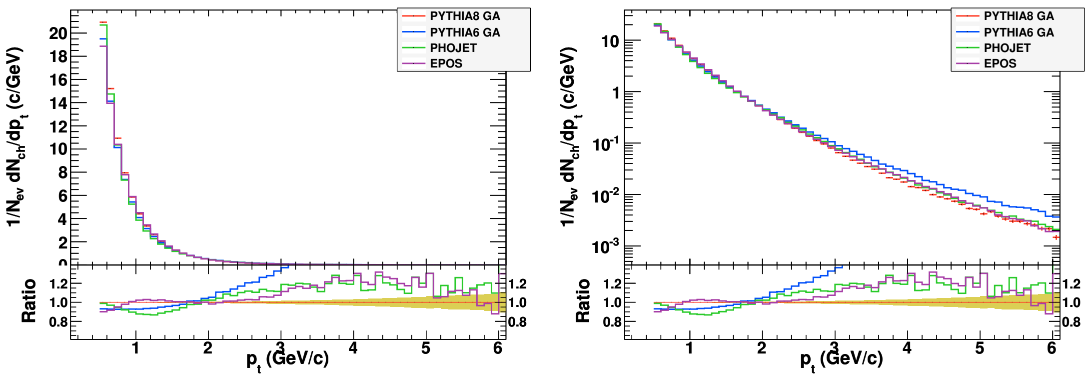


Figure 7.10.: Transverse momentum distribution for $|\eta| < 2.5$ at $\sqrt{s} = 2360$ GeV for events where MBTS_1 triggered. Linear scale is on the left, Logarithmic scale is on the right. The bottom plots show the ratio with respect to PYTHIA 8 distribution.

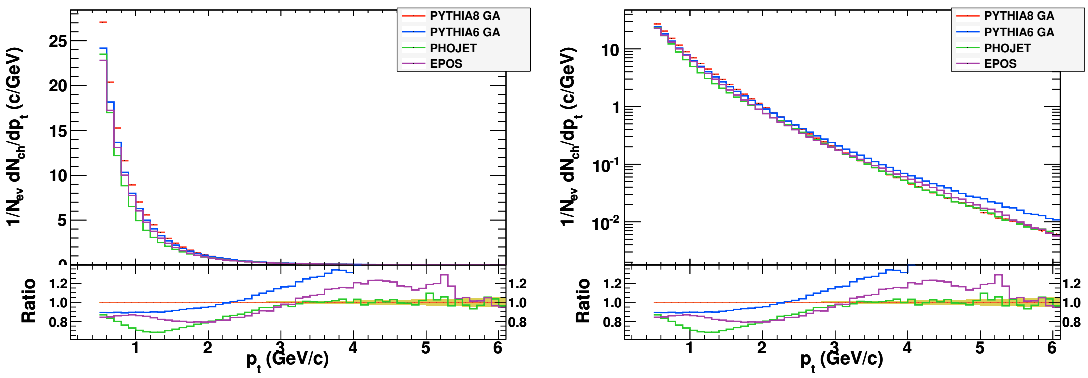


Figure 7.11.: Transverse momentum distribution for $|\eta| < 2.5$ at $\sqrt{s} = 7$ TeV for events where MBTS_1 triggered. Linear scale is on the left, Logarithmic scale is on the right. The bottom plots show the ratio with respect to PYTHIA 8 distribution.

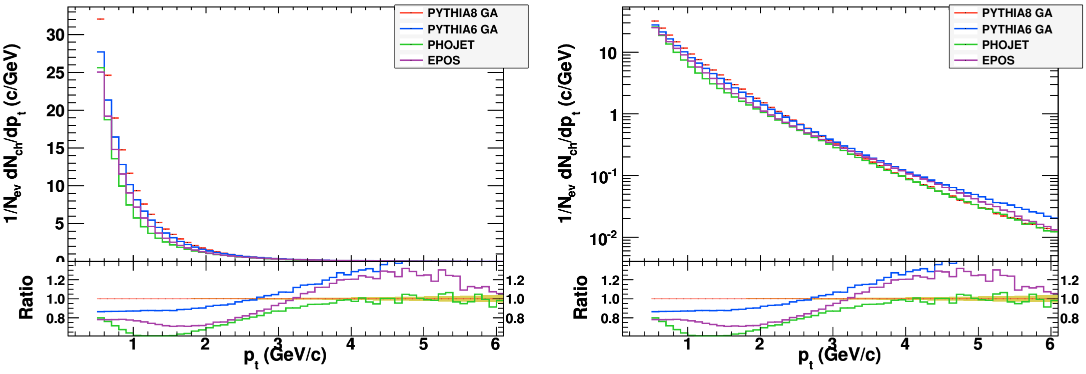


Figure 7.12.: Transverse momentum distribution for $|\eta| < 2.5$ at $\sqrt{s} = 14$ TeV for events where MBTS_1 triggered. Linear scale is on the left, Logarithmic scale is on the right. The bottom plots show the ratio with respect to PYTHIA 8 distribution.

smallest slope and PHOJET has the largest which can be attributed to its description of high multiplicity events.

7.4. Full Detector Simulation

Full detector simulation of the ATLAS is done using the experiment's software framework ATHENA [80]. Generator events are fed into the Geant4 [81] based simulation package to simulate particle-matter interactions, particle decays and charge depositions. Digitization package converts this information into signals of the detector electronics, creating pseudo-data. The pseudo-data is then reconstructed by the reconstruction algorithms for estimating the efficiencies, acceptances and preparing physics analyses.

For the production of the full simulation samples in this section the ATHENA version *15.5.4.10, Tier0* is used. Two database tags, *ConditionsTag* and *DetDescrVersion* defining the detector specific constants such as position of the detector elements, voltage settings etc. are taken as "OFLCOND-SIM-BS2T-00" and "ATLAS-GEO-08-00-02", respectively. Low-p_{T} track reconstruction is enabled for the determination of primary vertices. The remaining parameters are taken as defaults.

ATHENA does not have an interface for EPOS, thus an interface is implemented for generation of the EPOS samples. The default PHOJET interface in ATHENA uses PYTHIA 6.4.21. This is known to differ from the official PHOJET version, which is interfaced to PYTHIA 6.1.15. Thus PHOJET samples require retuning of the parameters for the new PYTHIA 6 interface. Figure 7.17 shows comparisons of $dN_{\mathrm{ch}}/d\eta$ and $dN_{\mathrm{ch}}/dp_{\mathrm{T}}$ distributions for PHOJET using PYTHIA 6.1.15 and 6.4.22 at $\sqrt{s} = 900$ GeV. Similar differences in distributions is observed at higher energies.

Event selection for the production of the plots required the existence of a primary vertex in the event. Primary vertex reconstruction in the ATLAS experiment is a complicated procedure and done using several different methods. The details of the primary

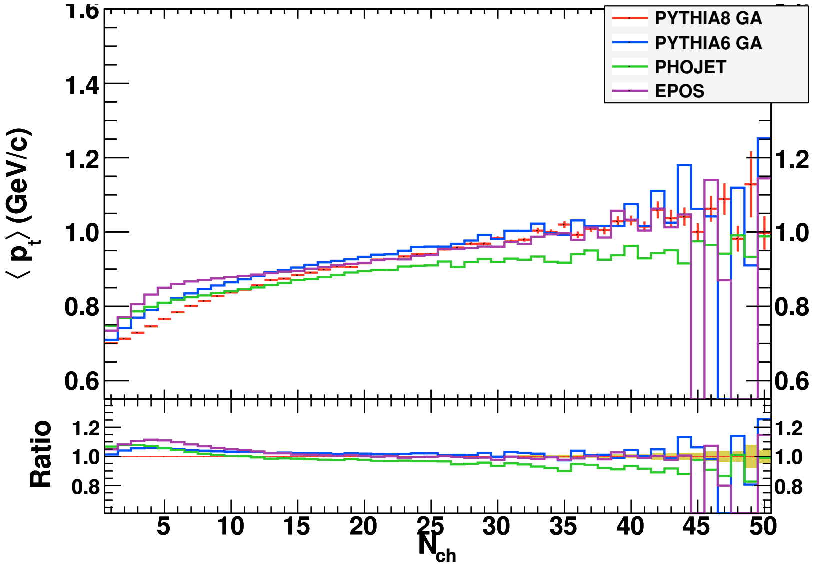


Figure 7.13.: Average p_T distribution for particles with $p_T > 0.5$ GeV at different N_{ch} in $|\eta| < 2.5$ at $\sqrt{s} = 900$ GeV for events where the MBTS_1 triggers. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

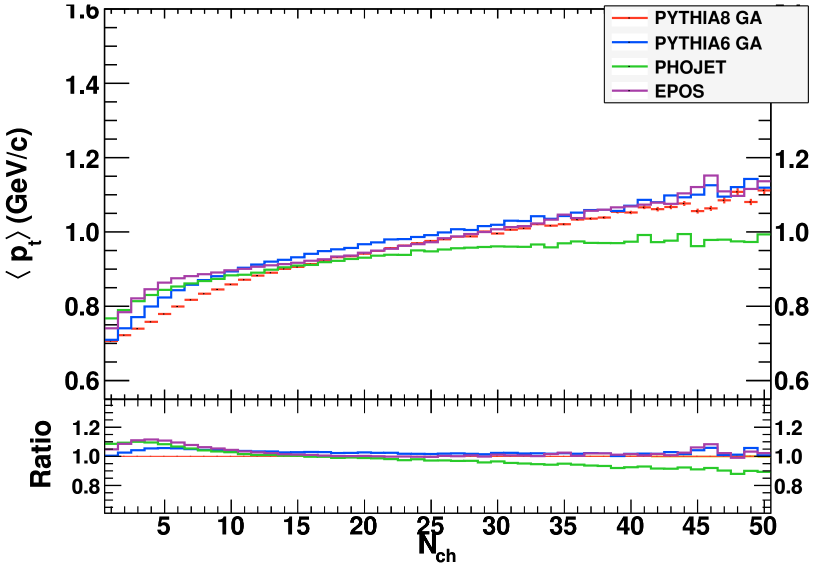


Figure 7.14.: Average p_T distribution for particles with $p_T > 0.5$ GeV at different N_{ch} in $|\eta| < 2.5$ at $\sqrt{s} = 2360$ GeV for events where the MBTS_1 triggers. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

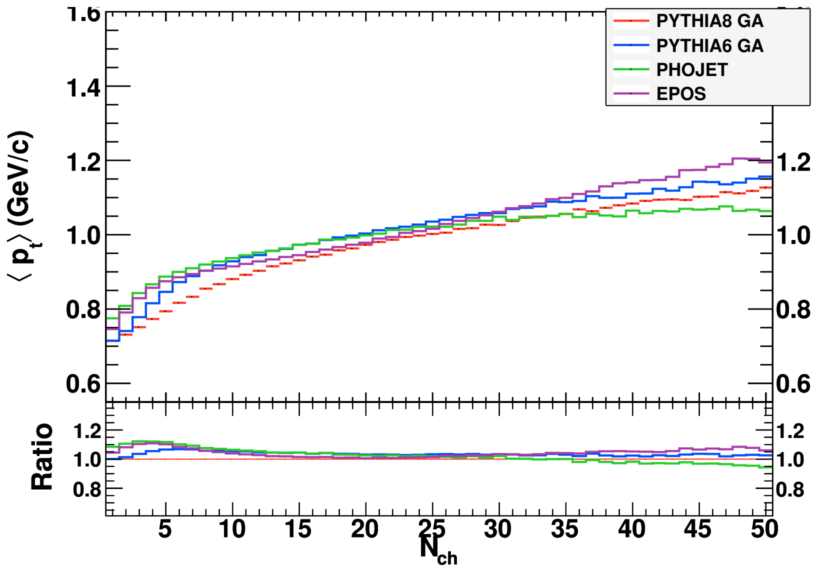


Figure 7.15.: Average p_{T} distribution for particles with $p_{\mathrm{T}} > 0.5$ GeV at different N_{ch} in $|\eta| < 2.5$ at $\sqrt{s} = 7$ TeV for events where the MBTS_1 triggers. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

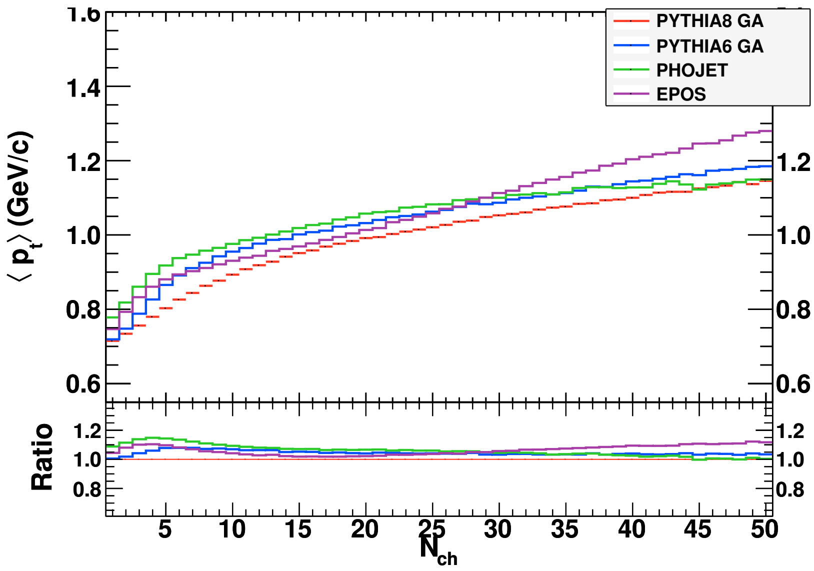


Figure 7.16.: Average p_{T} distribution for particles with $p_{\mathrm{T}} > 0.5$ GeV at different N_{ch} in $|\eta| < 2.5$ at $\sqrt{s} = 14$ TeV for events where the MBTS_1 triggers. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

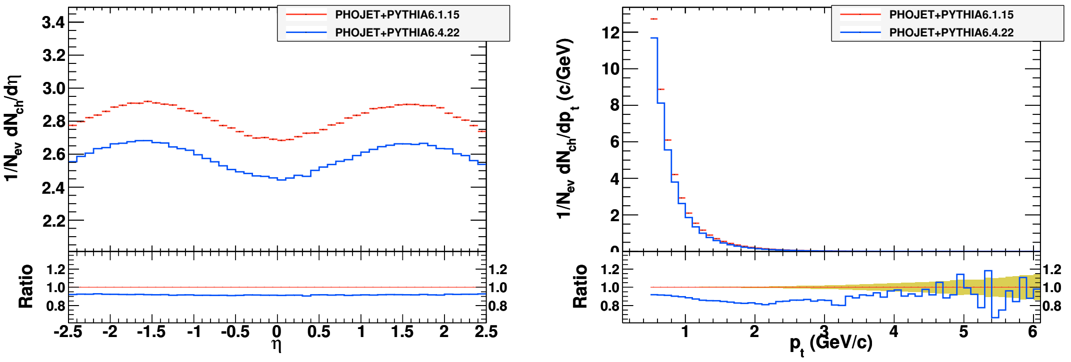


Figure 7.17.: Charged particle pseudorapidity and transverse momentum distributions of PHOJET interfaced to PYTHIA6.1.15 and PYTHIA6.4.22. Left and right plots show $dN_{\mathrm{ch}}/d\eta$ and $dN_{\mathrm{ch}}/dp_{\mathrm{T}}$ distributions at $\sqrt{s} = 900$ GeV, respectively.

vertex reconstruction can be found in [82] and the references therein. For the plots shown here, it was required that the primary vertex was reconstructed using at least three tracks, including those found by low-p_{T} reconstruction algorithms in the tracking software package.

Events are selected by requiring a *L1_MBTS_1* trigger, the existence of a primary vertex and at least one track satisfying the cuts listed in table 7.1. The L1_MBTS_1 efficiency is 100% for all generators at 900, 2360 and 7000 GeV except for very low multiplicity events, where it drops down to a minimum of 99.85% at 7000 GeV. The primary vertex finding efficiency at 900 GeV is shown in figure 7.18. It is better than 98% for events with more than one primary tracks at all three energies and has negligible dependence on the generator. For single primary track events, it is better than 70% for all studied center-of-mass energies.

Track Property	**cut**
p_{T}	≥ 0.5 GeV
$\lvert\eta\rvert$	≤ 2.5
Pixel Detector Hits	≥ 1
SCT Detector Hits	≥ 6
Track reconstruction algorithm	inside-out
$\lvert d_0\rvert$ wrt Primary Vertex	$\leq 1.5\,\mathrm{mm}$
$\lvert z_0 \sin(\theta)\rvert$ wrt Primary Vertex	$\leq 1.5\,\mathrm{mm}$

Table 7.1.: Primary track selection cuts.

Figures 7.19 to 7.21 show the charged particle multiplicity distribution of the selected tracks. At 900 GeV the distributions are close to each other. However at 2360 GeV the differences reach 15% and become as big as 30% at 7000 GeV. Agreement between the generator level distributions for $\sqrt{s} = 900$ GeV that are presented in chapter 6 and the

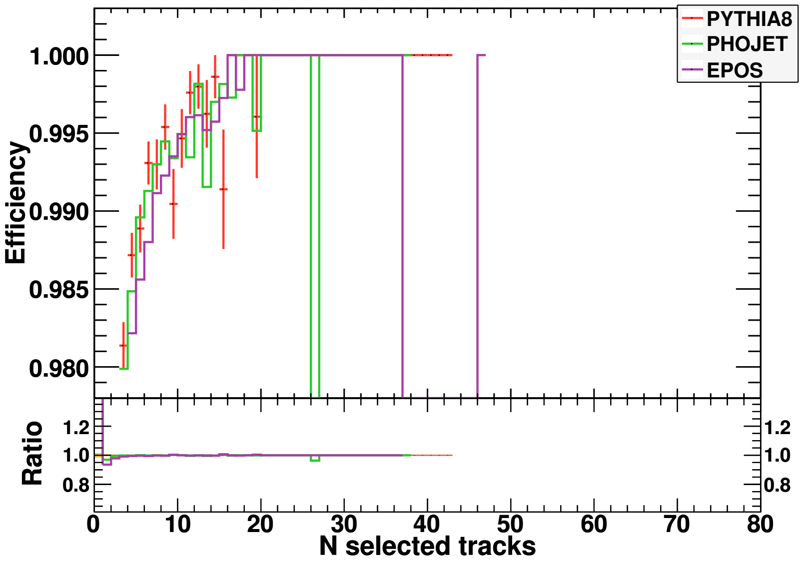


Figure 7.18.: Primary Vertex finding efficiency as a function of number of tracks satisfying the cuts in table 7.1 at $\sqrt{s} = 900$ GeV. The bottom plot shows the ratio with respect to the PYTHIA 8 distribution.

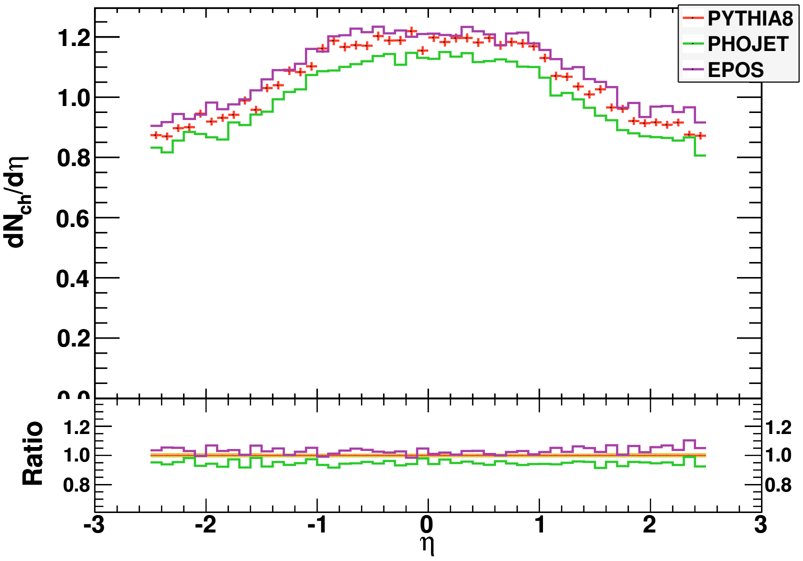


Figure 7.19.: Reconstructed $dN_{\text{ch}}/d\eta$ distribution at $\sqrt{s} = 900$ GeV. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

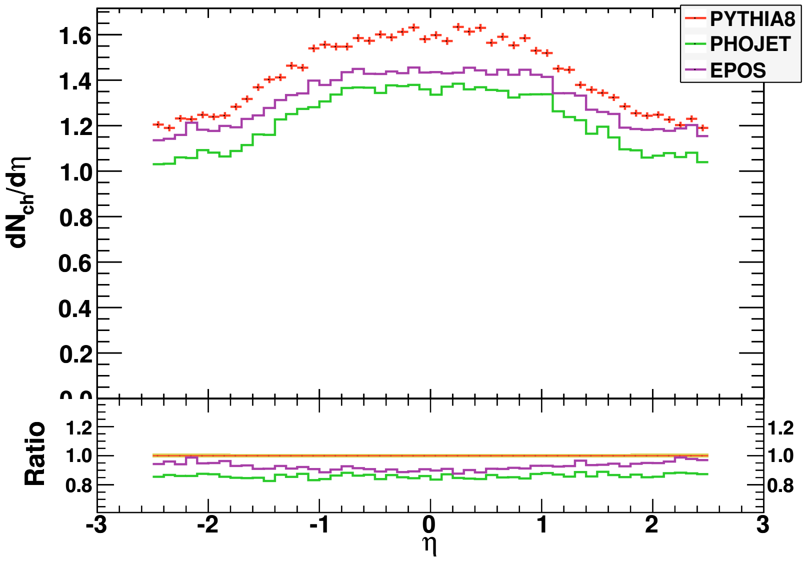


Figure 7.20.: Reconstructed $dN_{\mathrm{ch}}/d\eta$ distribution at $\sqrt{s} = 2360$ GeV. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

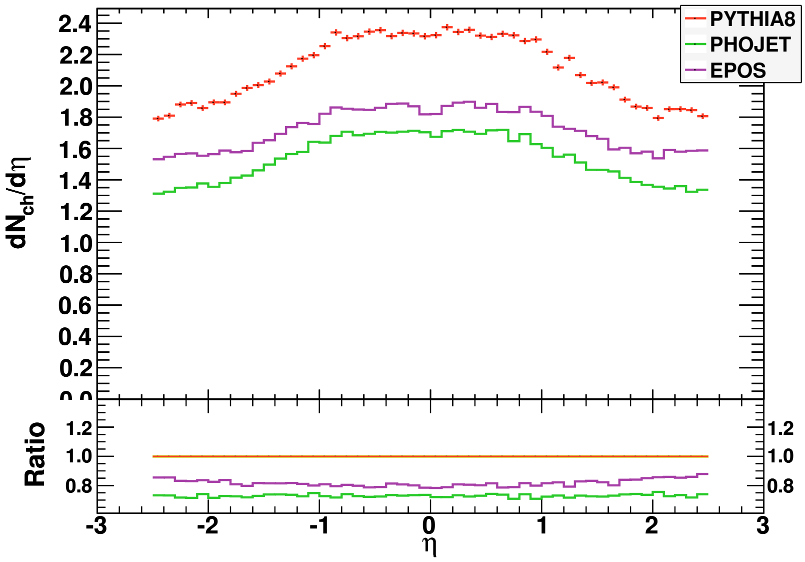


Figure 7.21.: Reconstructed $dN_{\mathrm{ch}}/d\eta$ distribution at $\sqrt{s} = 7$ TeV. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

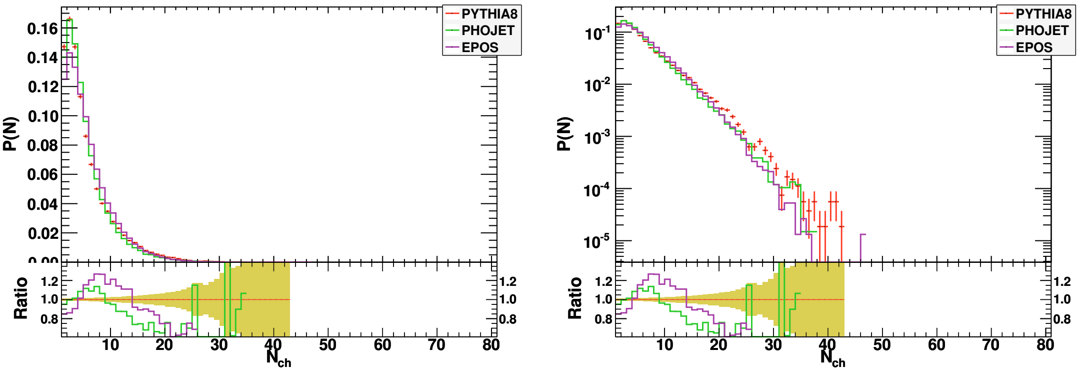


Figure 7.22.: Reconstructed charged particle multiplicity at $\sqrt{s} = 900$ GeV. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

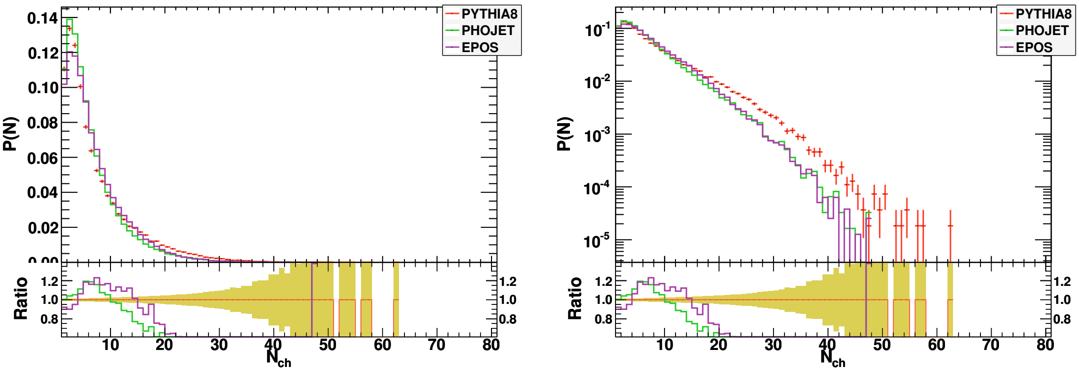


Figure 7.23.: Reconstructed charged particle multiplicity at $\sqrt{s} = 2360$ GeV. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

reconstructed distributions in this chapter are within 10% of each other in the region $|\eta| < 1.0$ but it gets as big as 25% outside. The large differences in PHOJET distributions are coming from the different PYTHIA interfaces. At higher energies the differences between generator level and reconstructed values get larger and show a dependence on the generator. Usually acceptances and inefficiencies are corrected by using various methods such as unfolding. However such methods should be applied carefully, since some dependence on the generators can be observed in the figures.

The track multiplicity plots shown in figures 7.22 to 7.24, the p_T plots in figures 7.25 to 7.27 and the average p_T distributions in figures 7.28 to 7.30 follow the trend observed with generator level plots. These plots also show some degree of generator dependency.

Comparing the generator level distributions in the previous sections with the full reconstruction plots shown in this section, one can see that the ATLAS reconstruction framework can reproduce the truth information very successfully especially in the barrel region. $dN_{ch}/d\eta$ plots show slightly degraded reconstruction efficiency for $1.0 < |\eta| < 2.5$.

Reconstructed distributions from all generators at $\sqrt{s} = 900$ GeV are similar to each

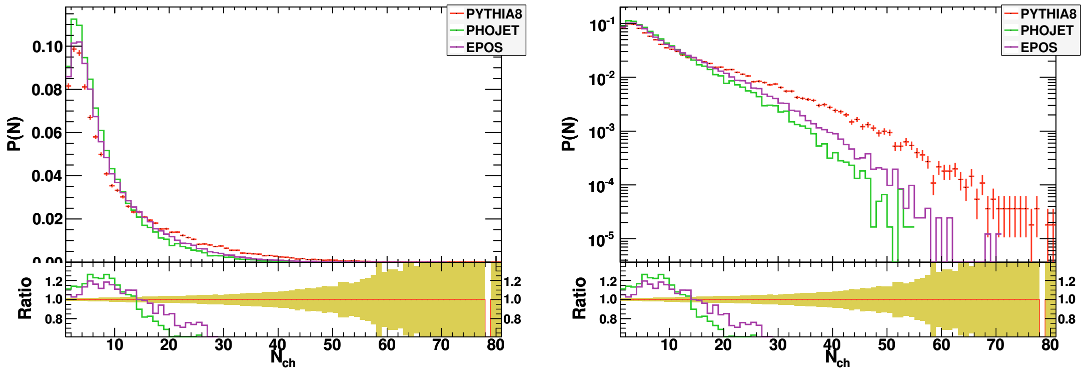


Figure 7.24.: Reconstructed charged particle multiplicity at $\sqrt{s} = 7$ TeV. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

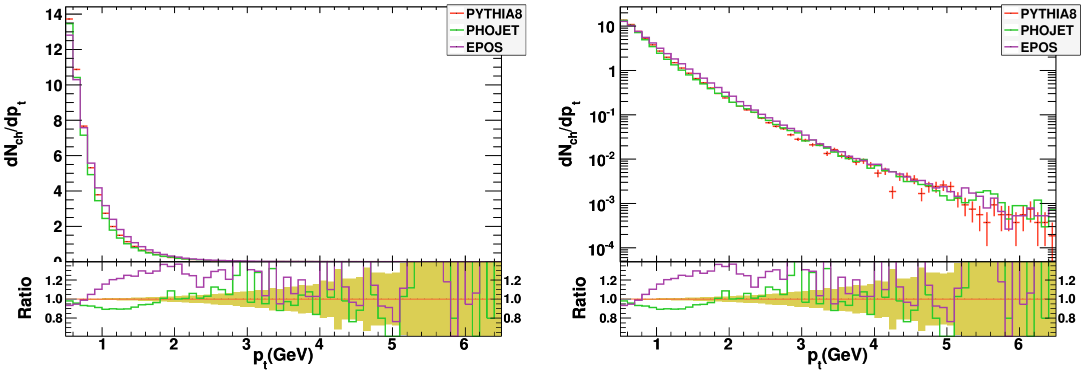


Figure 7.25.: Reconstructed p_T distribution of the tracks at $\sqrt{s} = 900$ GeV. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

other and a discrimination between the models is unlikely with the 900 GeV data. On the other hand, data collected at higher energies should provide enough discrimination.

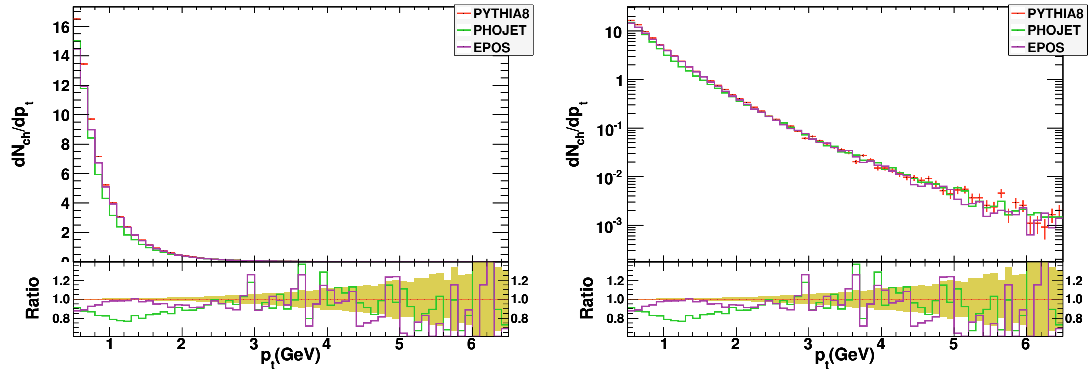


Figure 7.26.: Reconstructed p_T distribution of the tracks at $\sqrt{s} = 2360$ GeV. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

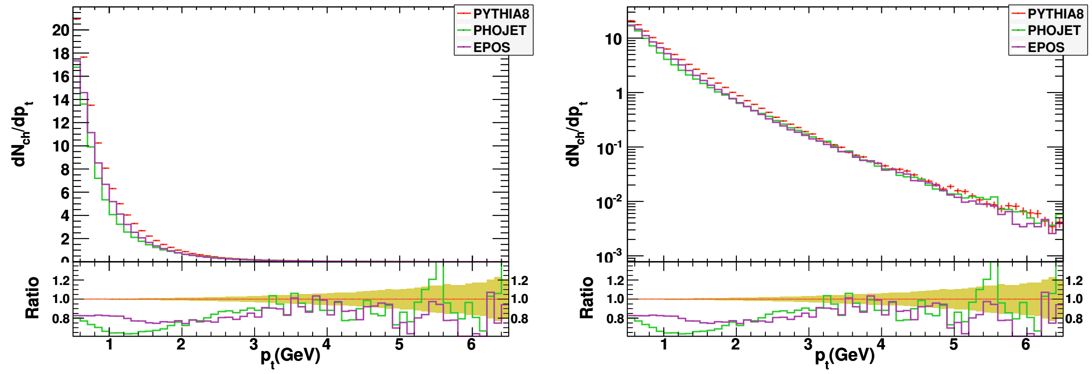


Figure 7.27.: Reconstructed p_T distribution of the tracks at $\sqrt{s} = 7$ TeV. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

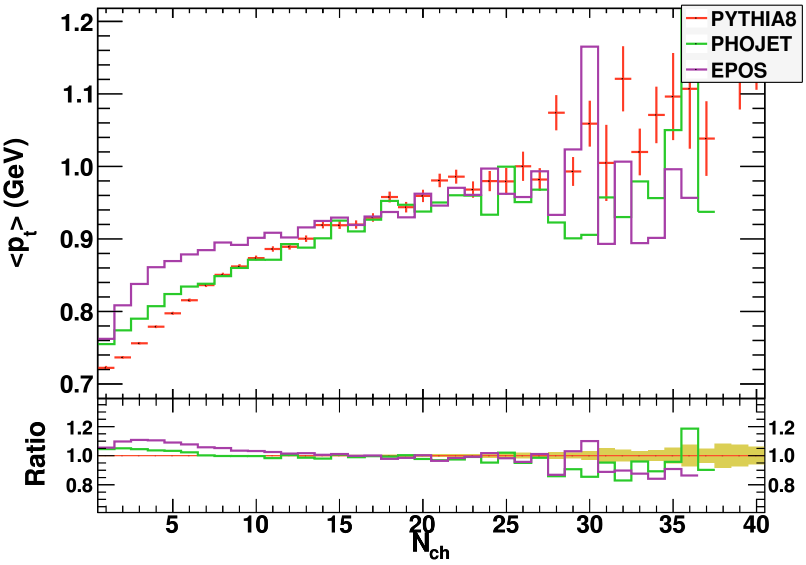


Figure 7.28.: Average p_T versus number of reconstructed tracks at $\sqrt{s} = 900$ GeV. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

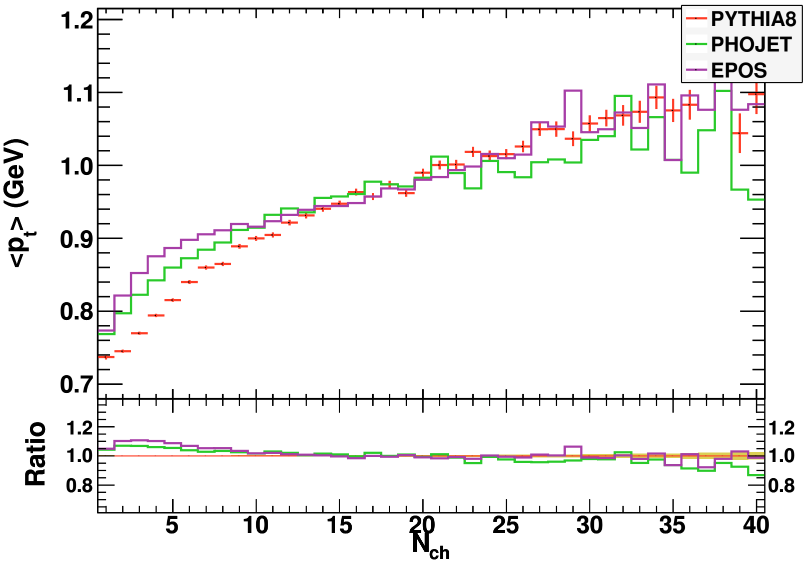


Figure 7.29.: Average p_T versus number of reconstructed tracks at $\sqrt{s} = 2360$ GeV. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

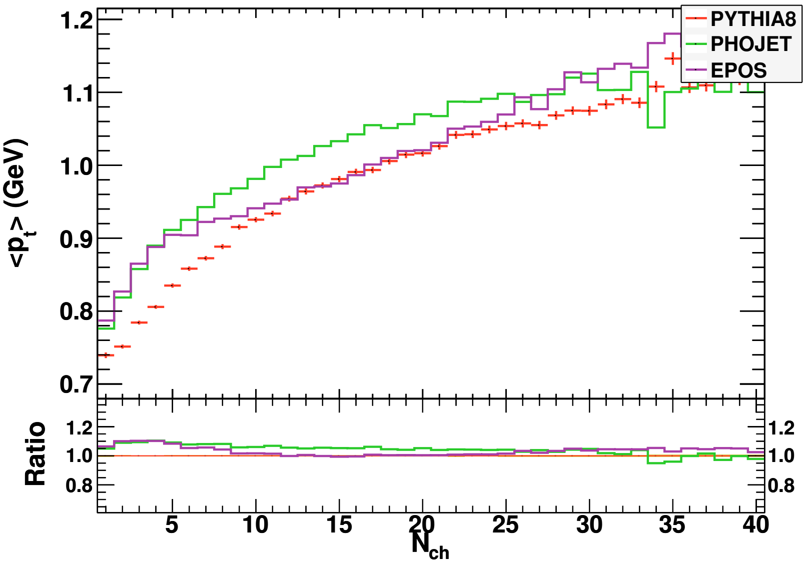


Figure 7.30.: Average p_T versus number of reconstructed tracks at $\sqrt{s} = 7$ TeV. The bottom plot shows the ratio with respect to PYTHIA 8 distribution.

8. Conclusions

Genetic Algorithms have been used in various fields of science since their invention in the 1960s. They found various applications in high energy physics as well. However, it is for the first time that the Genetic Algorithms are used for tuning complex Monte Carlo programs in a fully automatic way. It uses exact generator response to drive the tuning and the results are also exact. The most important drawback of the method is long running times which is a consequence of its use of exact MC response. However it is reduced in exchange for resources by distributing the evaluation of the MC response over several computers. Since nowadays it is rather easy to access cheap multi-core computers, even small groups can run tuning in relatively short times. However, MC tunes are usually used by collaborations which have access to large computing resources, making the resource problem negligible.

The Genetic Algorithm model and genetic operations implemented in this thesis are one the most basic types of Genetic Algorithms. It is possible to use advanced GA models and improve the efficiency of the framework. Thus this thesis also serves as a proof-of-principle and opens a way for further exploitation of the technique.

The tuning method described in this thesis is applied to PYTHIA 6 and 8. The tunes found by the method are comparable to or better than the tunes found by using other methods and can describe pre-LHC data equally well.The PYTHIA6 GA tune describes the $\sqrt{s} = 900$ GeV CMS $dN_{\mathrm{ch}}/d\eta$ distribution well, however, it underestimates the 2360 GeV data where PHOJET and EPOS is in better agreement.

The new diffractive model implemented in PYTHIA 8 improves its description of minimum bias data, especially for high multiplicity events.

The ATLAS Detector will measure pseudorapidity and transverse momentum of charged particles in the range $|\eta| < 2.5$ as low as $p_{\mathrm{T}} > 0.15$ GeV. Basic minimum bias observables such as $dN_{\mathrm{ch}}/d\eta$, charged particle multiplicity, transverse momentum and mean p_{T} versus event multiplicity distributions will be measured in this kinematic domain. Such distributions have been investigated using different soft interaction models implemented in EPOS, PHOJET and PYTHIA using GA tunes and found that they all have acceptable agreement with the existing data. Soft interactions produced by PYTHIA 8 using a new diffractive model and EPOS are studied in comparison with PYTHIA 6 and PHOJET, two most commonly used generators in minimum bias studies. Both EPOS and PYTHIA 8 are found to be better than PHOJET and PYTHIA 6. PHOJET has been found to fail to correctly describe some distributions and some indications of bugs have been observed. PHOJET distributions can be improved by using recent parton distribution functions.

EPOS is found to be successful in describing minimum bias data. The ATLAS experiment will benefit from it.

Full simulation studies using the ATLAS reconstruction framework show that recent data at $\sqrt{s} = 900$ GeV is not enough to provide information for favoring one model over another. However, data at high energies should provide enough discrimination power between the models. Total cross-section measurements, charged particle densities at $\eta = 0$, multiplicity distributions at different η ranges, $dN_{\mathrm{ch}}/d\eta$ distributions or average p_{T} distributions at $\sqrt{s} = 7$ TeV can provide enough distinction between the models.

The new charged particle pseudorapidity data from ATLAS and CMS can not be described with a single model and different models work better for either data set. This shows that although the models can be tuned for a specific kinematical domain, none of them successfuly describes it. This could create problems at higher energies. However experiments can give more feedback to theory by publishing their measurements corrected for detector and trigger effects with as little dependency on a model as possible. Also in order to constrain models it would be beneficial if experiments can publish their data in various kinematical domains such as $Ed^3\sigma/dp^3$ distributions at various η or N_{ch} intervals, $1/N_{\mathrm{event}} dN_{\mathrm{event}}/dN_{\mathrm{ch}}$ at various η intervals. In that way models can be constrained more efficiently and phasespace can be better studied.

Appendix

A. Operational Monitoring Display

Operational Monitoring Display (OMD) is prepared for providing a fast yet representative information of the operational status of the ATLAS TDAQ systems [83]. In order to achieve this it is designed in two parts, OMD Core and OMD GUI.

A.1. OMD Core

OMD Core is responsible of gathering the information from the Information Service (IS) servers, calculating statistics, creating histograms and providing the information for the GUI. It is composed of several components, IS Gatherer, Storage, Classifier, Configurator and Histogram Producer. It uses push-pull and signaling approaches together and runs in multiple threads. Figure A.1 shows a schematic of the data and control flow inside OMD Core.

A.1.1. IS Gatherer

IS Gatherer is the entry point of the information on the IS servers into the OMD. It uses the subscription mechanism of the IS to receive the information. Subscription mechanism sends the information to its subscribers when the information is updated. The IS Gatherer converts the IS Objects into internal data type *ISG Objects*, making the rest of the core independent of the different object types in TDAQ. It also calculates the standard deviations, averages and sums of the ISG Objects between updates. But the chaotic structure of TDAQ information flow, due to thousands of applications producing and sending information, makes it almost impossible to have a constant update interval. ISGatherer produces a signal for starting the calculations when the same object is encountered a second time and most of the existing objects are updated for the given interval. After the calculations it signals the Storage, notifying it about the updates.

A.1.2. Storage

Storage pulls the updates from the IS Gatherer and puts them into a custom depth cyclic history buffer. It notifies the Classifier, Histogram Producer and GUI about the updates. It keeps the history of the objects and runs as a separate thread.

A.1.3. Classifier

Classifier groups the objects using the patterns in their names with the help of the regular expressions. Several regular expressions can be applied consecutively, each generating

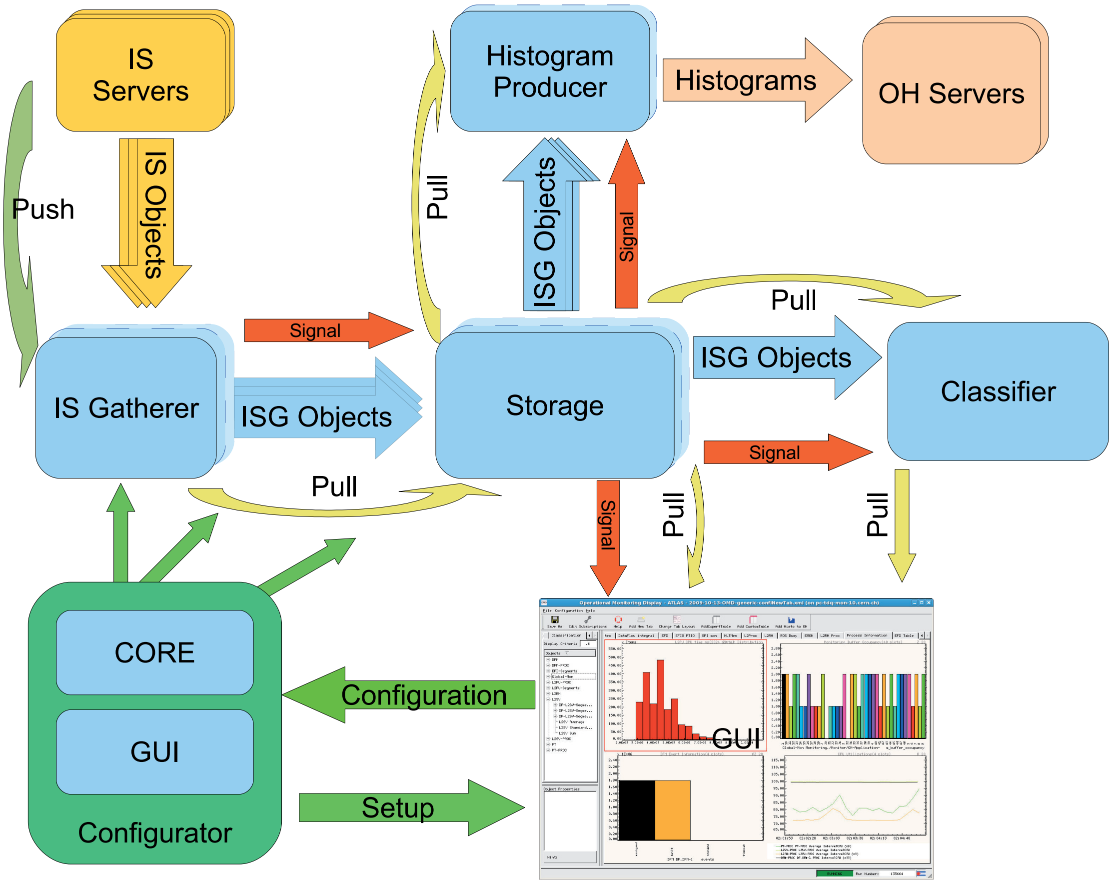


Figure A.1.: Data and control flow in the OMD core. Different components of the core is represented as the boxes. Multiple boxes correspond the objects which runs as separate threads and support multiple copies.

another level in the classification tree. At each update Classifier checks for new objects and updates the trees as needed. Moreover it calculates averages, sums and standard deviations at each level. Since it uses the patterns in the names and regular expressions in classification, it is mostly independent of the configuration of the underlying partition.

A.1.4. Histogram Producer

Histogram producer creates histograms from ISG Objects. It supports three different types of histograms, Frequency, Bar Chart and Bar Chart 2D. Frequency type provides the histograms of a field of an object. At each object update the updated field value is filled into the histogram. Bar Chart type encodes the value of a field of selected objects into a one dimensional histogram. X-bins represent the objects and bin content the value of the field. At each update histogram contents are replaced with new values. Bar Chart 2D type is a two dimensional version of Bar Chart type, containing all fields of the selected objects as the second dimension. At each update contents of the histogram is replaced. Figure A.2 contains a diagram showing the information used in creation of histograms.

Histogram Producer creates a bridge for OH based applications to access IS information. Also histograms can be saved and used in post-mortem analysis of the TDAQ.

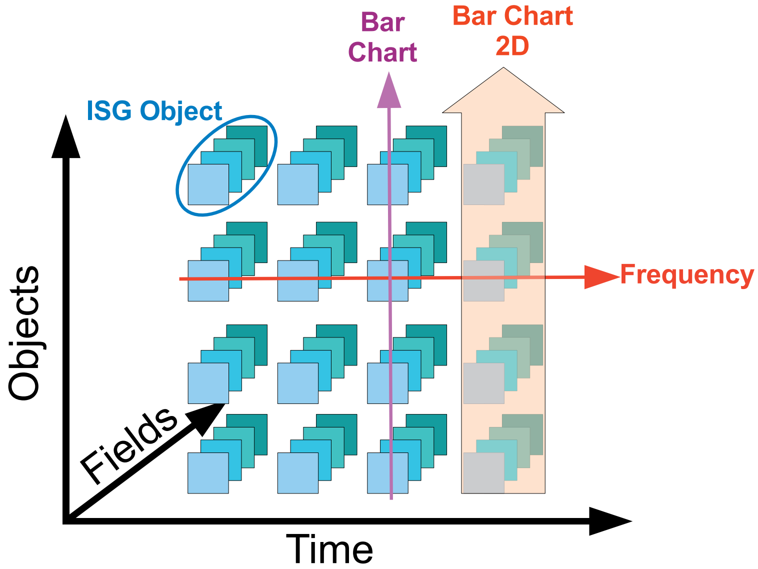


Figure A.2.: Schematics showing the contents of the histograms created by the Histogram Producer. The horizontal axis represents the time, vertical axis represents the objects and depth represents the fields of an object. Arrows show the information used in the histograms generated by Histogram Producer.

-c <configuration-file>	use previously saved “configuration-file”
-p <partition-name>	monitor “partition-name”
-d <configuration-directory>	show files in “configuration-directory” in start-up
-o <old-format-configuration-file>	try to convert “old-format-configuration-file” to new format
-C <configuration-files> <>..	try to merge given “configuration-files”
-h history-size	history depth size of each object.

Table A.1.: OMD command-line options.

A.1.5. Configurator

ISG Configurator manages the configuration of the GUI and the core. It keeps the configuration information in internal configuration data type. It handles the IS subscriptions of the IS Gatherer, instantiates the Storage, Classifier, Histogram Producer and makes the connections between them. It saves the configuration information in XML format into the configuration file. It can also combine several files into one file.

A.2. OMD GUI

OMD GUI is the main application where the information collected in the Core is presented and the user interaction takes place. It can be started from the command-line with the “OMD” command with the options listed on table A.1. If no option is given, it pops-up a window to select the partition and browse for existing configuration files. If no configuration file is selected, it starts with an empty window. After selection of a partition it shows the main window. An example of the main window with different types of plots can be seen in figure A.3.

The main window has a toolbar containing several buttons for various actions and its configuration. From left to right these buttons are “Save As”, “Edit Subscriptions”, “Help”, “Add New Tab”, “Change Tab Layout”, “Add Expert Table”, “Add Custom Table”, “Add Histo to OH”.

Briefly explanation of the buttons are; the *Save As* button saves current configuration to a new file, *Edit Subscriptions* button brings up the *Subscription Editor* window and the *Help* button shows online help. The *Add New Tab* creates a new tab, *Change Tab Layout* changes the layout and the number of plots in current tab. The *Add Expert Table* creates a table showing the numerical value of the selected field of a selected object and *Add Custom Table* creates a table where the user can select the fields to be displayed. Finally the *Add Histo to OH* brings up the Histogram Configuration window.

OMD GUI can display different types of plots in many tabs. New tabs can be added from the *Add New Tab* toolbar button. Each tab can contain up to 10 plots in $M \times N$ format which can be dynamically changed. The *Change Tab Layout* brings up a small window to change the number and layout of the plots in the tab. Each tab can be

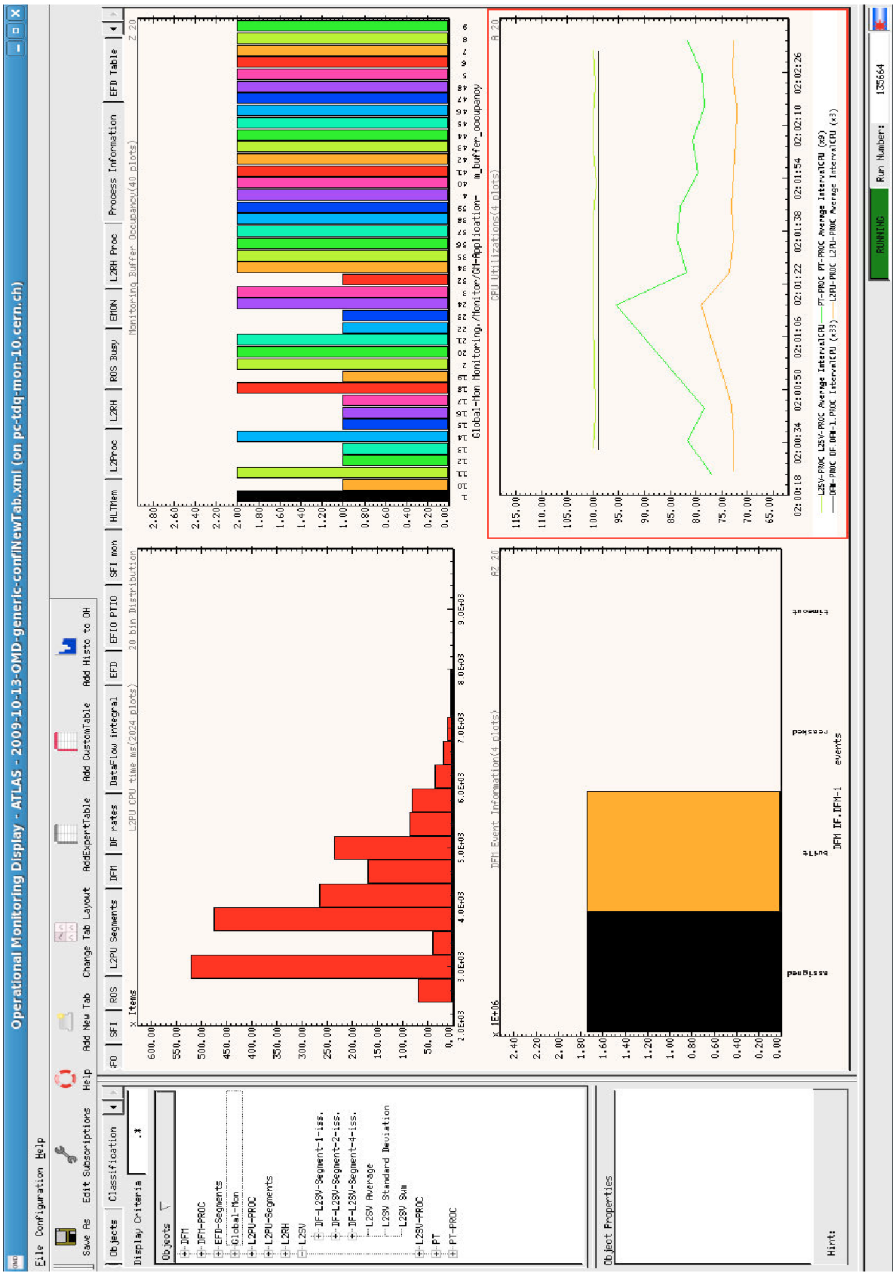
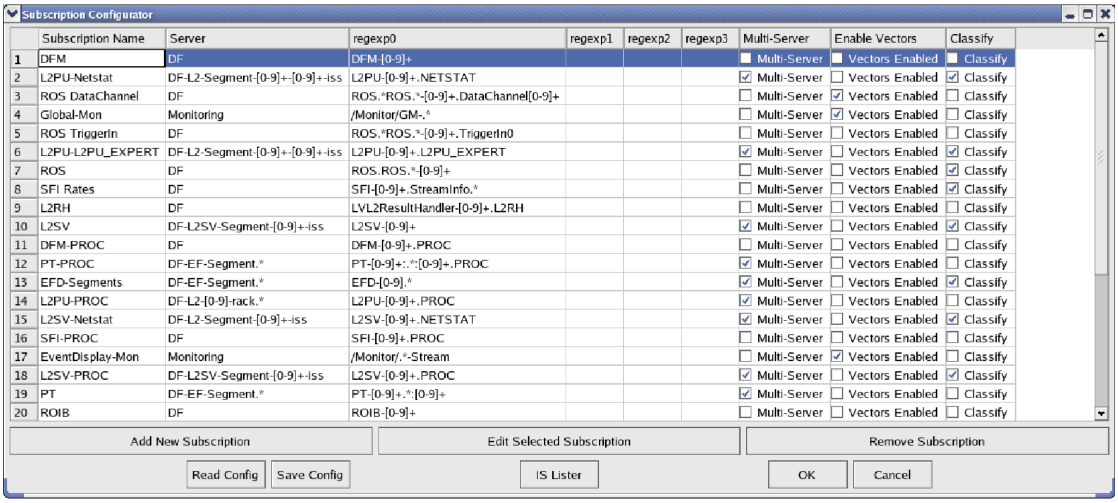


Figure A.3.: OMD Main Window. On the left, classified list of objects can be seen. On the right, data is presented in four different styles. Top-left plot is “Distribution” style, top-right is “Bar Chart”, bottom-left is “Histogram” and bottom-right is “Time-Series” style. Status-bar shows current run status, run number and other information.

detached from the main window using middle-click menu, becoming a separate window. Tabs can also be renamed or closed through the middle-click menu.

A.2.1. Subscription Editor

In order for OMD to operation, it should subscribe to IS servers for information. Subscriptions are configured through *Subscription Editor* window which is brought up by the Subscription Editor button in the toolbar and shown in figure A.4.

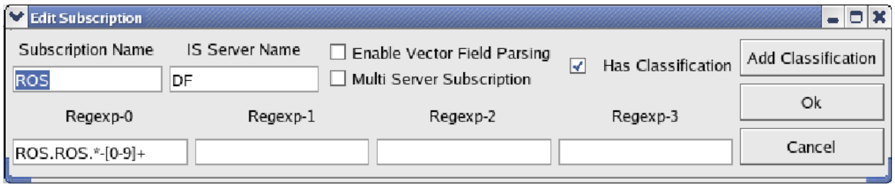


	Subscription Name	Server	regexp0	regexp1	regexp2	regexp3	Multi-Server	Enable Vectors	Classify
1	DFM	DF	DFM-[0-9]+				☐ Multi-Server	☐ Vectors Enabled	☐ Classify
2	L2PU-Netstat	DF-L2-Segment-[0-9]+-[0-9]+-iss	L2PU-[0-9]+.NETSTAT				☑ Multi-Server	☐ Vectors Enabled	☑ Classify
3	ROS DataChannel	DF	ROS.*ROS.*-[0-9]+.DataChannel[0-9]+				☐ Multi-Server	☑ Vectors Enabled	☐ Classify
4	Global-Mon	Monitoring	/Monitor/GM-.*				☐ Multi-Server	☑ Vectors Enabled	☐ Classify
5	ROS TriggerIn	DF	ROS.*ROS.*-[0-9]+.TriggerIn0				☐ Multi-Server	☐ Vectors Enabled	☐ Classify
6	L2PU-L2PU_EXPERT	DF-L2-Segment-[0-9]+-[0-9]+-iss	L2PU-[0-9]+.L2PU_EXPERT				☑ Multi-Server	☐ Vectors Enabled	☑ Classify
7	ROS	DF	ROS.ROS.*-[0-9]+				☐ Multi-Server	☐ Vectors Enabled	☑ Classify
8	SFI Rates	DF	SFI-[0-9]+.StreamInfo.*				☐ Multi-Server	☐ Vectors Enabled	☑ Classify
9	L2RH	DF	LVL2ResultHandler-[0-9]+.L2RH				☐ Multi-Server	☐ Vectors Enabled	☐ Classify
10	L2SV	DF-L2SV-Segment-[0-9]+-iss	L2SV-[0-9]+				☑ Multi-Server	☐ Vectors Enabled	☑ Classify
11	DFM-PROC	DF	DFM-[0-9]+.PROC				☐ Multi-Server	☐ Vectors Enabled	☐ Classify
12	PT-PROC	DF-EF-Segment.*	PT-[0-9]+:.*:[0-9]+.PROC				☑ Multi-Server	☐ Vectors Enabled	☐ Classify
13	EFD-Segments	DF-EF-Segment.*	EFD-[0-9].*				☑ Multi-Server	☐ Vectors Enabled	☑ Classify
14	L2PU-PROC	DF-L2-[0-9]-rack.*	L2PU-[0-9]+.PROC				☑ Multi-Server	☐ Vectors Enabled	☐ Classify
15	L2SV-Netstat	DF-L2-Segment-[0-9]+-iss	L2SV-[0-9]+.NETSTAT				☑ Multi-Server	☐ Vectors Enabled	☑ Classify
16	SFI-PROC	DF	SFI-[0-9]+.PROC				☐ Multi-Server	☐ Vectors Enabled	☐ Classify
17	EventDisplay-Mon	Monitoring	/Monitor/.*-Stream				☐ Multi-Server	☑ Vectors Enabled	☐ Classify
18	L2SV-PROC	DF-L2SV-Segment-[0-9]+-iss	L2SV-[0-9]+.PROC				☑ Multi-Server	☐ Vectors Enabled	☑ Classify
19	PT	DF-EF-Segment.*	PT-[0-9]+.*:[0-9]+				☑ Multi-Server	☐ Vectors Enabled	☐ Classify
20	ROIB	DF	ROIB-[0-9]+				☐ Multi-Server	☐ Vectors Enabled	☐ Classify

Figure A.4.: OMD Subscription Editor window.

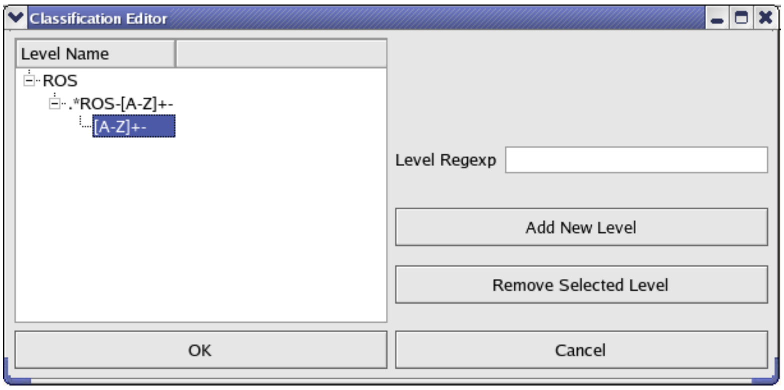
New subscriptions can be added by clicking to “Add New Subscription” button. Each subscription composed of several fields and can be edited through “Edit Subscription” window which is shown in figure A.5. **Subscription Name** is the unique name of the subscription and it can not be empty. **IS Server Name** is the name of the IS server to be subscribed. It can not be empty but it can be a regular expression if **Multi Server Subscription** check-box is selected. In that case core will subscribe to all servers that match the regular expression and collect the objects under the same subscription name. The fields **Regexp-0** to **Regexp-3** are the fields for object-matching regular expressions used in subscriptions. At least one of them must have a regular expression. Several regular expressions increase the control over selection procedure and makes it possible to select object sets which can not be selected with single regular expression. Any object matching one of these regular expressions on the server or servers in **IS Server Name** must be of the same type and will be grouped under the **Subscription Name** and listed on the main window. **Enable Vector Field Parsing** check-box enables the parsing of vectors in the objects. By default they are disabled in order to increase the performance.

Objects selected in the subscription can be classified by adding a classification configuration to subscription. **Add Classification** button brings up the “Classification Window” shown in figure A.6. Classification editor shows the subscription name as its root. New levels of regular expressions can be added by typing it in **Level Regexp** box



Subscription Name: ROS
IS Server Name: DF
☐ Enable Vector Field Parsing
☐ Multi Server Subscription
☑ Has Classification
Regexp-0: ROS.ROS.*-[0-9]+
Regexp-1:
Regexp-2:
Regexp-3:

Figure A.5.: Edit Subscription window. Showing several fields that are forming a subscription.



Level Name
- ROS
  - .*ROS-[A-Z]+-
    - [A-Z]+-

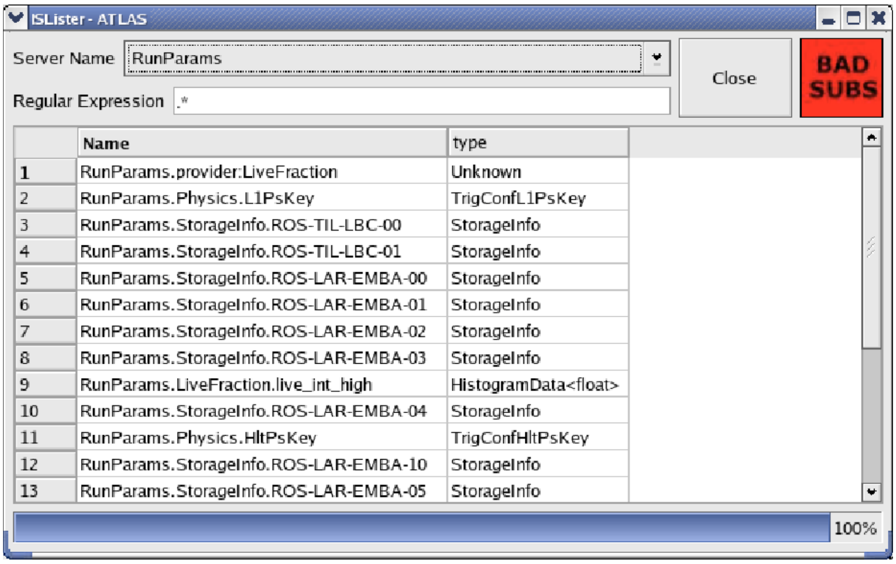
Level Regexp:

Figure A.6.: Classification Editor window showing the classification regular expressions as a list with different levels. Top level shows the subscription name. Subsequent levels show the regular expressions used in creation of respective levels in the classification tree.

and clicking **Add New Level** button. Regular expressions are applied to object names, one after another following their levels. For each matching regular expression a new item in the respective level is created and objects sharing the same string placed under that item. The process is repeated until all objects are assigned to an item in the last level or no objects matched by the current level. In this way a tree structure is formed and individual statistics are calculated for each level and for each item. These trees are shown under **Classification** tab in main window. If a subscription has classification configuration then **Has Classification** and **Classify** check-boxes are checked.

Edit Selected Subscription button in the “Subscription Configurator” window brings up the “Add New Subscription” window, filled with the values of selected subscription. **Read Config** button reads the subscription configuration from the configuration file and **Save Config** writes it back.

IS Lister button brings up a “IS Lister” helper widget shown in figure A.7. IS Lister can check whether give regular expression is valid, whether all matched objects are of the same type and show all objects matching it. It shows red, yellow and green signs showing the results of checks. Users can test their regular expressions with IS Lister and see whether it selects the desired objects.



Server Name: RunParams

Regular Expression: .*

	Name	type
1	RunParams.provider:LiveFraction	Unknown
2	RunParams.Physics.L1PsKey	TrigConfL1PsKey
3	RunParams.StorageInfo.ROS-TIL-LBC-00	StorageInfo
4	RunParams.StorageInfo.ROS-TIL-LBC-01	StorageInfo
5	RunParams.StorageInfo.ROS-LAR-EMBA-00	StorageInfo
6	RunParams.StorageInfo.ROS-LAR-EMBA-01	StorageInfo
7	RunParams.StorageInfo.ROS-LAR-EMBA-02	StorageInfo
8	RunParams.StorageInfo.ROS-LAR-EMBA-03	StorageInfo
9	RunParams.LiveFraction.live_int_high	HistogramData<float>
10	RunParams.StorageInfo.ROS-LAR-EMBA-04	StorageInfo
11	RunParams.Physics.HltPsKey	TrigConfHltPsKey
12	RunParams.StorageInfo.ROS-LAR-EMBA-10	StorageInfo
13	RunParams.StorageInfo.ROS-LAR-EMBA-05	StorageInfo

Figure A.7.: IS Lister helper widget. Lists objects matching given regular expression in selected IS server. It displays red, yellow and green style indications about the regular expression.

A.2.2. Plots and Tables

OMD displays the information in the various forms of plots and tables. Figure A.3 shows examples of available plot types. All plots can be interchanged on the fly and each have their own internal custom depth history buffer. Plots are configured through right-click menu shown in figure A.8. Parts of right-click menu display different options depending on the plot type.

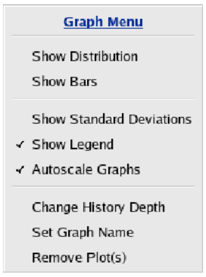
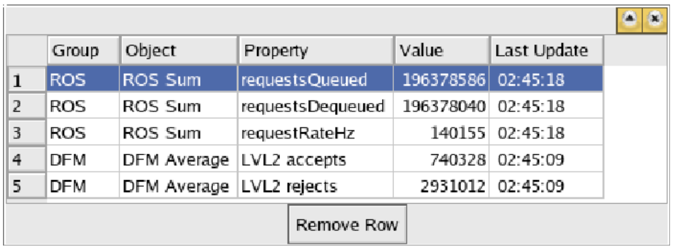


Figure A.8.: Right-Click Menu.

Time series plots present the evolution of the values over time. It supports automatic scaling of plots, multiplying the small valued curves with an integer factor and drawing the curve close to highest valued one. In that way axis range can be kept small, displaying the details of the curves. Multiplicative factors of curves are displayed next to their labels in legend on the bottom of plot.



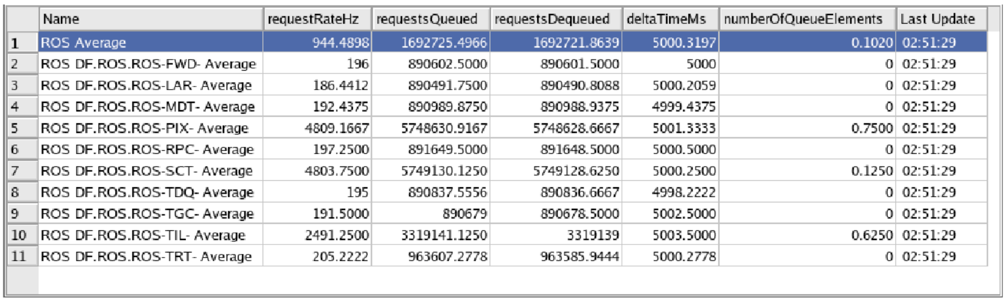
	Group	Object	Property	Value	Last Update
1	ROS	ROS Sum	requestsQueued	196378586	02:45:18
2	ROS	ROS Sum	requestsDequeued	196378040	02:45:18
3	ROS	ROS Sum	requestRateHz	140155	02:45:18
4	DFM	DFM Average	LVL2 accepts	740328	02:45:09
5	DFM	DFM Average	LVL2 rejects	2931012	02:45:09

Figure A.9.: Expert Table example.

Bar chart plots display the latest update of the value as bars. Instead of latest update, it is also possible to display the cumulative values, incrementing the bar with each update. Bar charts are very useful to spot outlying values. They also support auto-scaling, giving a measure of difference in case of large differences.

Values can be displayed in Distribution plots as well. Binning of the distribution can be changed from right-click menu and plot changes its axis dynamically to make sure all object displayed properly.

Numeric values are displayed in one of the two types of tables. “Expert Table”, shown in figure A.9, has fixed columns displaying the subscription name, object name, field of the object, the last value of the field and last update time. Each row can display only one field, but Expert Table can display many rows from different types of objects. Only one instance of Expert Table is allowed at any time. The other type, “Custom Table”, is shown in figure A.10 and can display several fields at the same time. Custom Tables are created with “Custom Table Creator” widget shown in figure A.11. **Table Name** is the unique name for the table. **Subscription Type** pull-down menu lists the available subscriptions containing objects. Once a subscription is selected **Property** pull-down menu is filled with the field names of the objects in the subscription. User can add any these fields in any order into the table with **Add Column** button. Generated Custom Table will display selected fields in columns in a row together with the object name and last update time. There can be many instances of Custom Table at any time however a Custom Table can display only the objects from the subscription that is used to construct it. Custom tables also have right-click menus.



	Name	requestRateHz	requestsQueued	requestsDequeued	deltaTimeMs	numberOfQueueElements	Last Update
1	ROS Average	944.4898	1692725.4966	1692721.8639	5000.3197	0.1020	02:51:29
2	ROS DF.ROS.ROS-FWD- Average	196	890602.5000	890601.5000	5000	0	02:51:29
3	ROS DF.ROS.ROS-LAR- Average	186.4412	890491.7500	890490.8088	5000.2059	0	02:51:29
4	ROS DF.ROS.ROS-MDT- Average	192.4375	890989.8750	890988.9375	4999.4375	0	02:51:29
5	ROS DF.ROS.ROS-PIX- Average	4809.1667	5748630.9167	5748628.6667	5001.3333	0.7500	02:51:29
6	ROS DF.ROS.ROS-RPC- Average	197.2500	891649.5000	891648.5000	5000.5000	0	02:51:29
7	ROS DF.ROS.ROS-SCT- Average	4803.7500	5749130.1250	5749128.6250	5000.2500	0.1250	02:51:29
8	ROS DF.ROS.ROS-TDQ- Average	195	890837.5556	890836.6667	4998.2222	0	02:51:29
9	ROS DF.ROS.ROS-TGC- Average	191.5000	890679	890678.5000	5002.5000	0	02:51:29
10	ROS DF.ROS.ROS-TIL- Average	2491.2500	3319141.1250	3319139	5003.5000	0.6250	02:51:29
11	ROS DF.ROS.ROS-TRT- Average	205.2222	963607.2778	963585.9444	5000.2778	0	02:51:29

Figure A.10.: A Custom Table example.

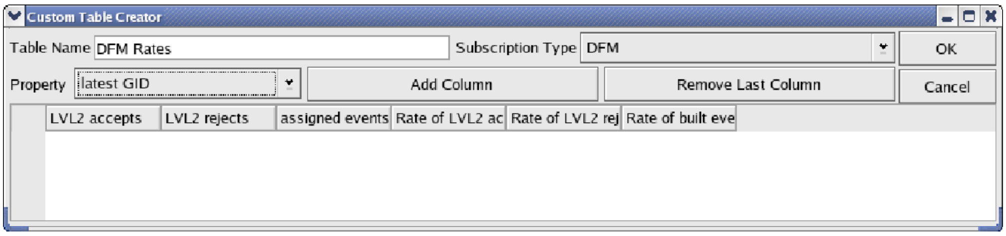


Table Name DFM Rates
Subscription Type DFM
Property latest GID

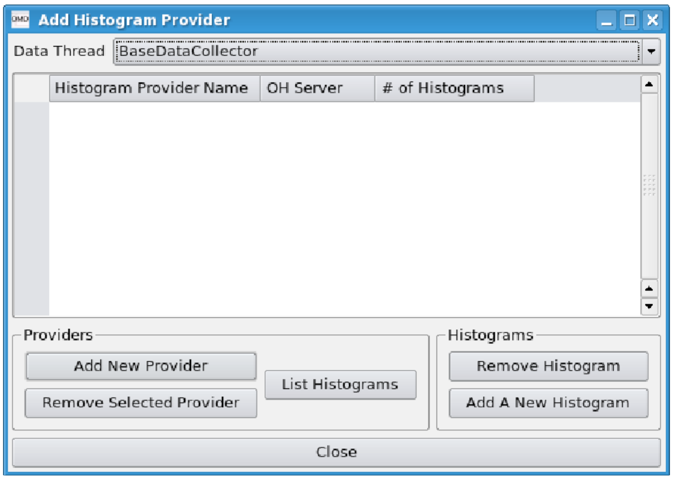
Figure A.11.: Custom Table Creator.

A.2.3. Histogram Configuration

Add Histo to OH button on the toolbar of the main window brings up "Add Histogram Producer" widget as shown in figure A.12. It lists available histogram producers and number of histograms they are producing. Each provider publishes histograms to a single OH server. **Add New Provider** button creates a new provider by asking a unique name and target OH server. Each provider runs as a separate thread. **Add A New Histogram** button brings up "Add Histogram" window. It supports three different type of histogram styles and dynamically changes to depending on the selected style. Figure A.13 shows layouts for Histogram and Bar Chart styles. **Histogram Name** is the unique name of the histogram in the histogramming server. **Object Type** pull-down menu lists available subscriptions containing objects. **Object Property** lists the fields of objects in the selected subscription. For "Bar Chart 2D" style histograms, Object Property menu is disabled since a "Bar Chart 2D" style histogram contains all fields of all objects. **X-Axis Label** and **Y-Axis Label** are the labels of the histogram axes. "Do Not Include Statistics in Histograms" check-box in Bar Chart and Bar Chart 2D styles controls whether calculated statistics objects included in histogram or not. Histogram style histograms contain fields for setting the number of bins, and the boundaries of histogram. **Object** pull-down menu provides the option for selecting either one of statistics objects or one of the objects in the subscription. **Select Object** button lists the objects in the subscription. Figure A.14 shows an example of each histogram style. Histogram style is common histogram. Bar Chart style encodes all objects in selected subscription into the histogram. Bins represent the objects and the content of the bin is the last value of the selected field. Bar Chart 2D is the two dimensional version of Bar Chart style. X-bins represents the objects, y-bins represents the fields and the bin content is the last value of the respective field.

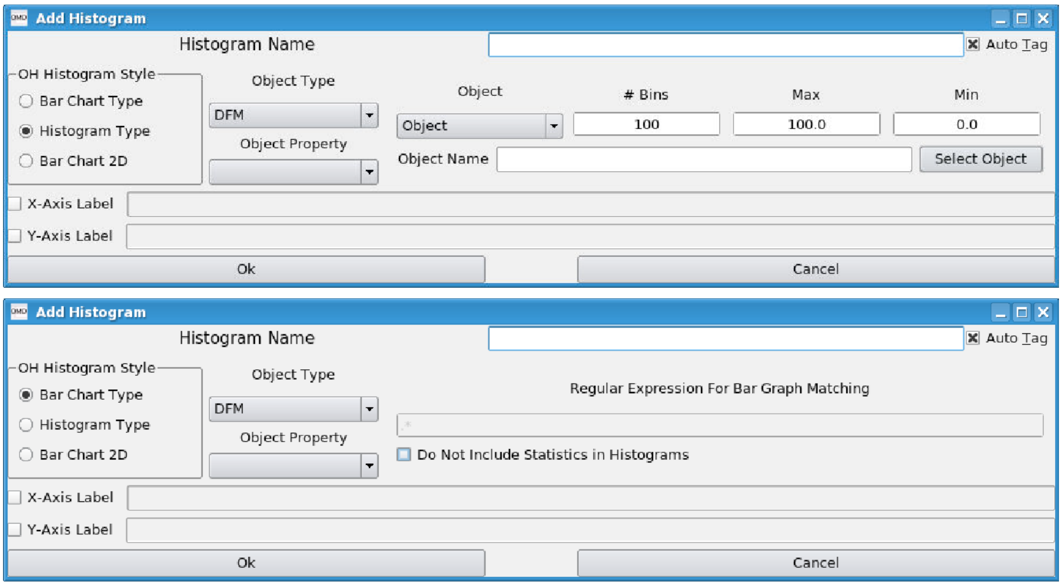
A.2.4. Alerts

OMD can support basic checks against thresholds and display messages. Alerts can be added through "Alerts" widget which is accessible from "Configuration" menu in main window. Figure A.15 shows the Alerts widget listing configured alerts. **Add Alert** button displays "Add Alarm" window as shown in figure A.16. Users add new alerts using Add Alarm window. **Alarm Name** is the unique name of the alarm. **Type** lists available subscriptions for which the alarm is going to be attached to. **Object** pull-



Data Thread: BaseDataCollector

Figure A.12.: Add Histogram Producer widget. It shows the list of available histogram producers and the number of histograms they are producing.



Histogram Name: ☒ Auto Tag

OH Histogram Style: ○ Bar Chart Type ● Histogram Type ○ Bar Chart 2D

Object Type: DFM

Object: Object

Bins: 100

Max: 100.0

Min: 0.0

☐ X-Axis Label

☐ Y-Axis Label

Histogram Name: ☒ Auto Tag

OH Histogram Style: ● Bar Chart Type ○ Histogram Type ○ Bar Chart 2D

Object Type: DFM

☐ Do Not Include Statistics in Histograms

☐ X-Axis Label

☐ Y-Axis Label

Figure A.13.: Add Histogram Widget. Upper plot shows layout for Histogram type and bottom plot shows layout Bar Chart type. Object Property pull-down menu in Bar Chart style is disabled in Bar Chart 2D style.

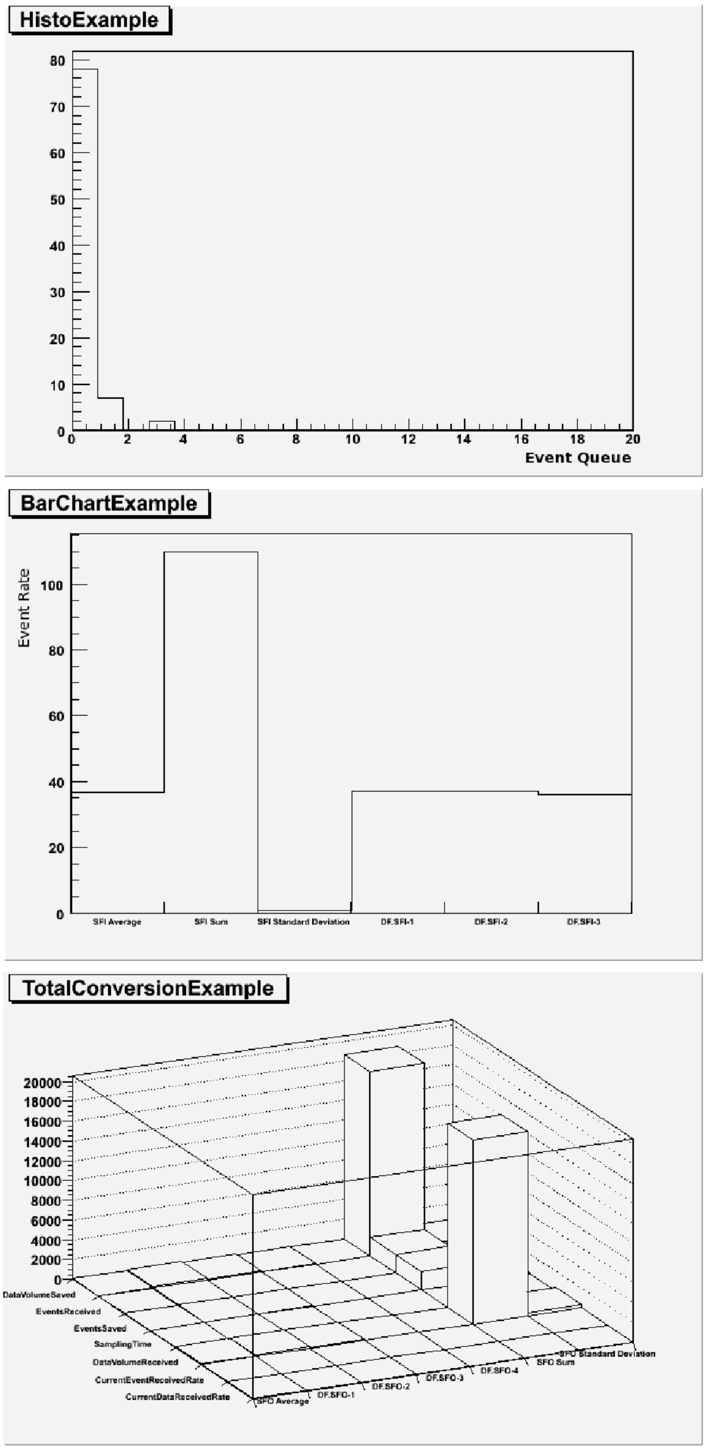
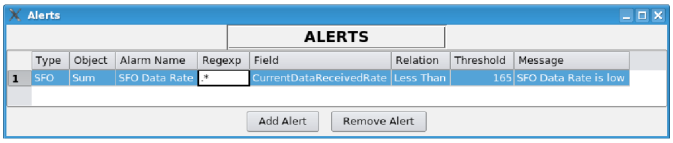
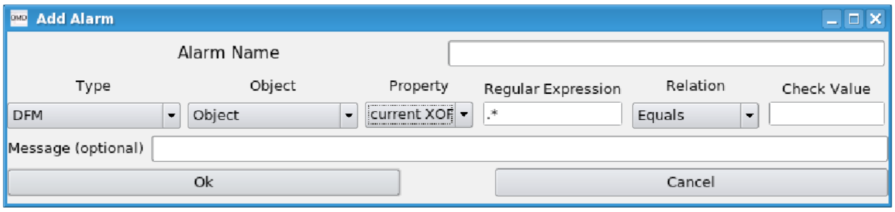


Figure A.14.: OMD histogram styles. Top plot is “Histogram” style, middle plot is “Bar Chart” and bottom plot is “Bar Chart 2D”.



	Type	Object	Alarm Name	Regexp	Field	Relation	Threshold	Message
1	SFO	Sum	SFO Data Rate	.*	CurrentDataReceivedRate	Less Than	165	SFO Data Rate is low

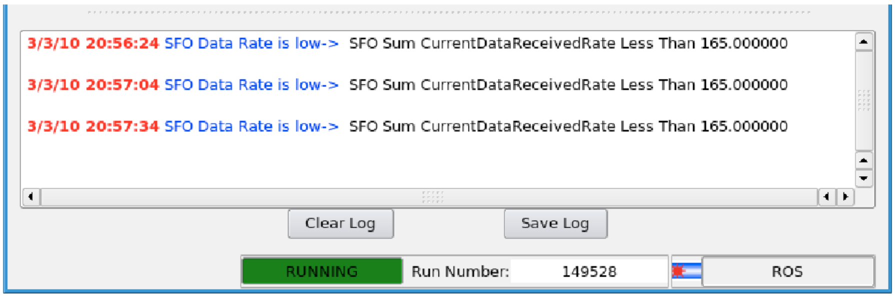
Figure A.15.: Alerts Widget. It lists available alerts.



Alarm Name:
Type: DFM
Object: Object
Property: current XOF
Regular Expression: .*
Relation: Equals
Check Value:
Message (optional):

Figure A.16.: Add Alarm window.

down menu for selecting whether the alarm is checked against the statistics objects of the subscription or regular objects. **Property** lists available fields of the objects in the subscription. **Regular Expression** is the regular expression for selecting a subset of the objects. It is disabled if alarm is checked against statistics objects. **Relation** is the logical operator to be used for testing. There are four operators available, namely *equal*, *not equal*,*less than* and *greater than*. **Check Value** is the threshold that the value of the field is compared to. Users can add custom message to alert using **Message** text-box. A message, including configured message, the object name and time, will be displayed in the Log Pane when the relation between the threshold and the selected field of the object which matches the regular expression. Figure A.17 shows the log pane. The "Log Pane" is closed by default to leave more space for plot area. It can be opened by using the handle above the status-bar. Alerts are not stored in the configuration file and valid only throughout the lifetime of the process.



```
3/3/10 20:56:24 SFO Data Rate is low-> SFO Sum CurrentDataReceivedRate Less Than 165.000000

3/3/10 20:57:04 SFO Data Rate is low-> SFO Sum CurrentDataReceivedRate Less Than 165.000000

3/3/10 20:57:34 SFO Data Rate is low-> SFO Sum CurrentDataReceivedRate Less Than 165.000000
```

Run Number: 149528

Figure A.17.: Log Pane showing alert messages.

B. Monte-Carlo Comparison Plots

This appendix contains data-MC comparison plots for the data sets that are described in chapter 5. The details of the analyses are given in section 5.1. Since PHOJET and EPOS can not produce Drell-Yan events, only PYTHIA 6 and 8 are used for Z-boson p_{T} shape distribution from DØ and underlying event and $e^+ - e^-$ pair p_{T} distribution analyses.

Figures B.1 to B.8 show particle multiplicity distributions from UA5 at $\sqrt{s} = 200$ GeV and $\sqrt{s} = 900$ GeV in different pseudorapidity intervals. Figures B.9 to B.11 show the Drell-Yan Z-boson p_{T} measurements from DØ. Figures B.12 to B.14 show normalized cross-sections and underlying event measurements from CDF at $\sqrt{s} = 1800$ GeV and $\sqrt{s} = 1960$ GeV.

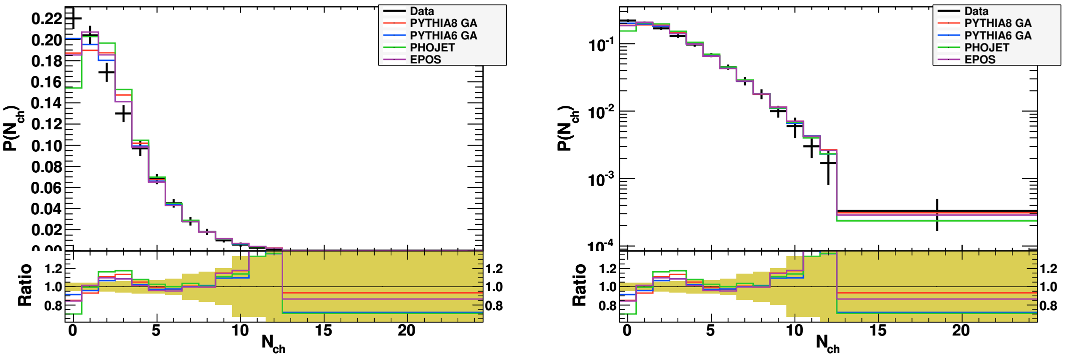


Figure B.1.: UA5 N_{ch} distribution in $|\eta| < 0.5$ at $\sqrt{s} = 200$ GeV. Right plot is linear and left plot is logarithmic scale [66].

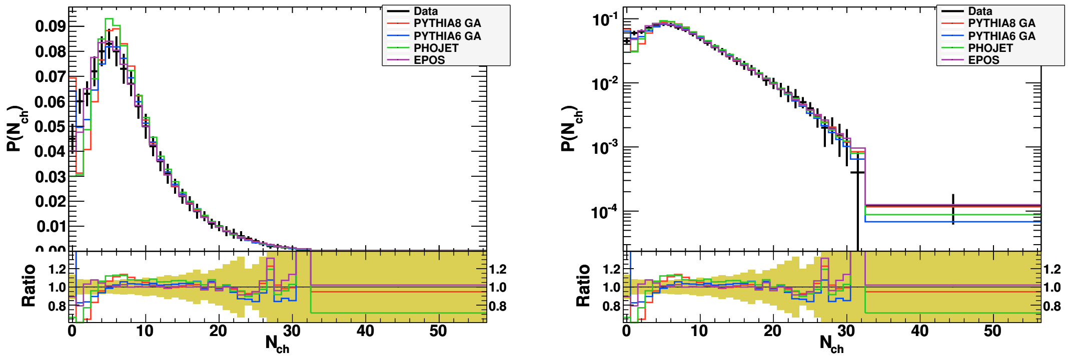


Figure B.2.: UA5 N_{ch} distribution in $|\eta| < 1.5$ at $\sqrt{s} = 200$ GeV. Right plot is linear and left plot is logarithmic scale [66].

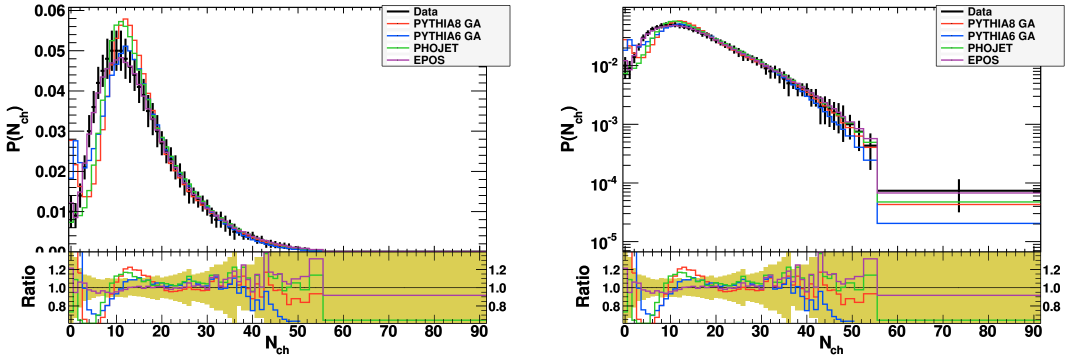


Figure B.3.: UA5 N_{ch} distribution in $|\eta| < 3.0$ at $\sqrt{s} = 200$ GeV. Right plot is linear and left plot is logarithmic scale [66].

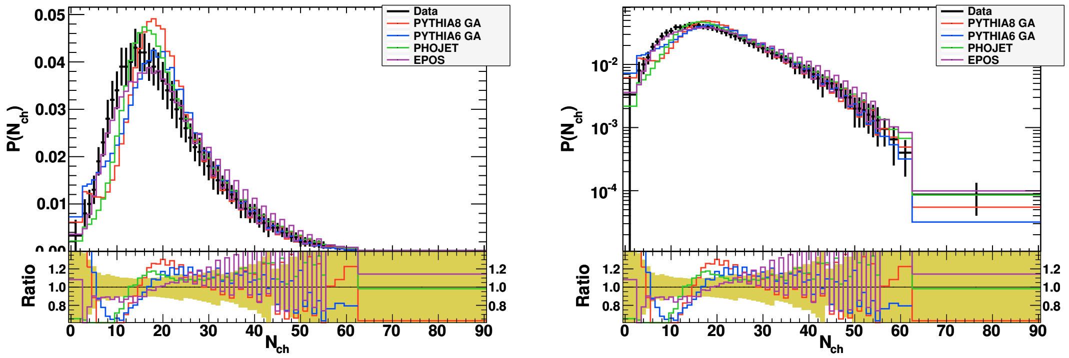


Figure B.4.: UA5 N_{ch} distribution in $|\eta| < 5.0$ at $\sqrt{s} = 200$ GeV. Right plot is linear and left plot is logarithmic scale [66].

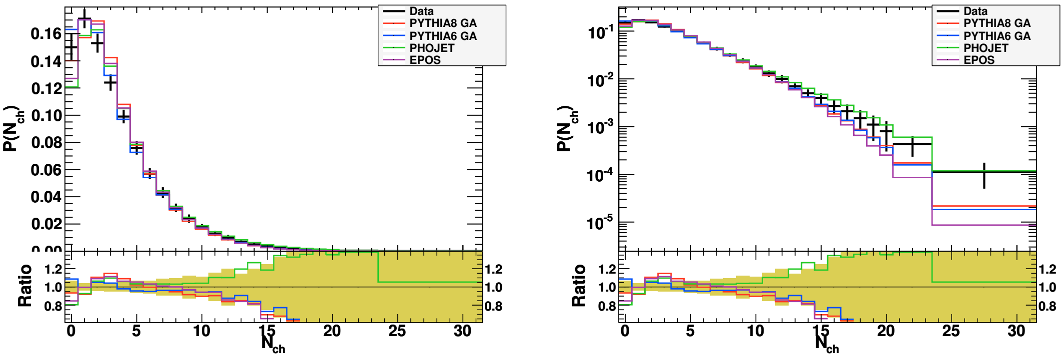


Figure B.5.: UA5 N_{ch} distribution in $|\eta| < 0.5$ at $\sqrt{s} = 900$ GeV. Right plot is linear and left plot is logarithmic scale [66].

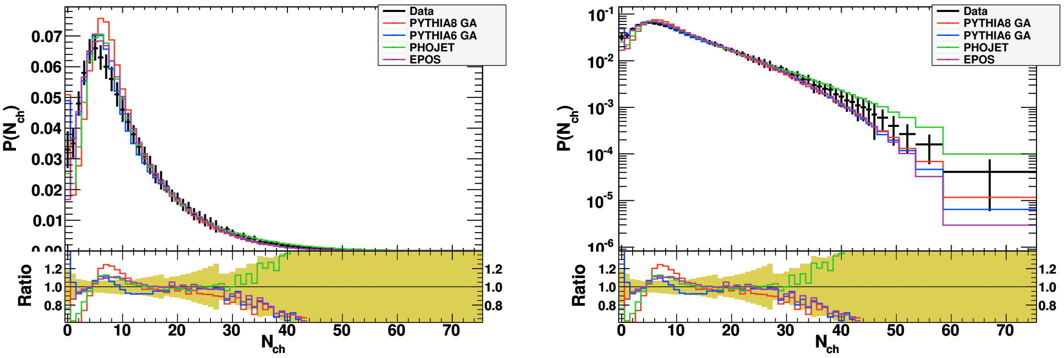


Figure B.6.: UA5 N_{ch} distribution in $|\eta| < 1.5$ at $\sqrt{s} = 900$ GeV. Right plot is linear and left plot is logarithmic scale [66].

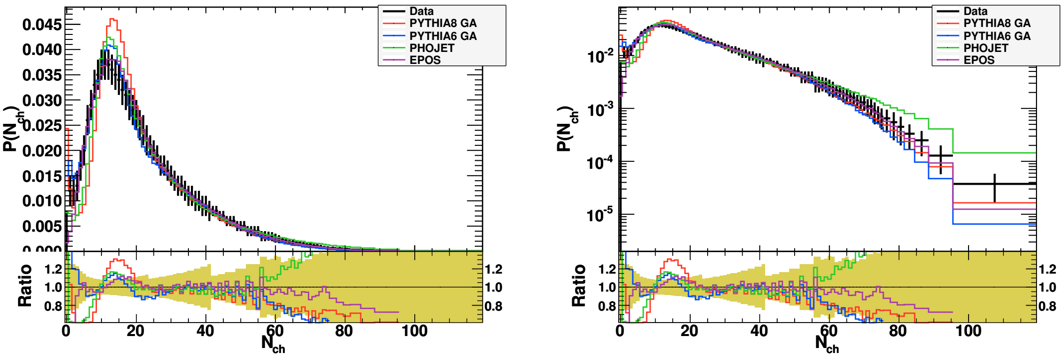


Figure B.7.: UA5 N_{ch} distribution in $|\eta| < 3.0$ at $\sqrt{s} = 900$ GeV. Right plot is linear and left plot is logarithmic scale [66].

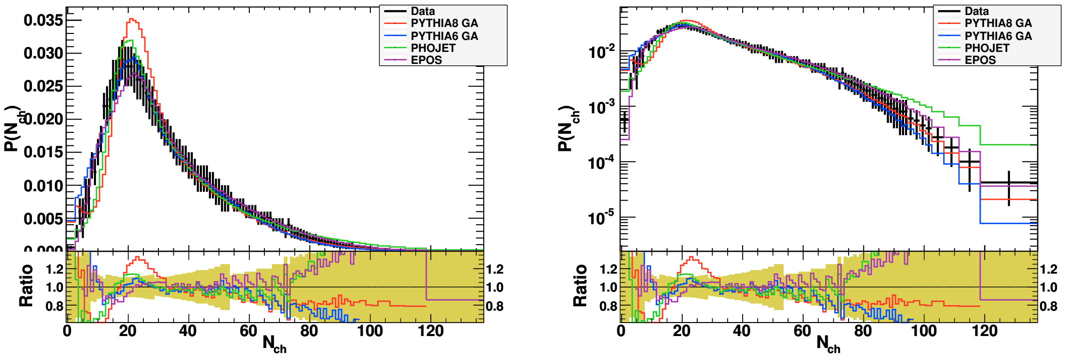


Figure B.8.: UA5 N_{ch} distribution in $|\eta| < 5.0$ at $\sqrt{s} = 900$ GeV. Right plot is linear and left plot is logarithmic scale [66].

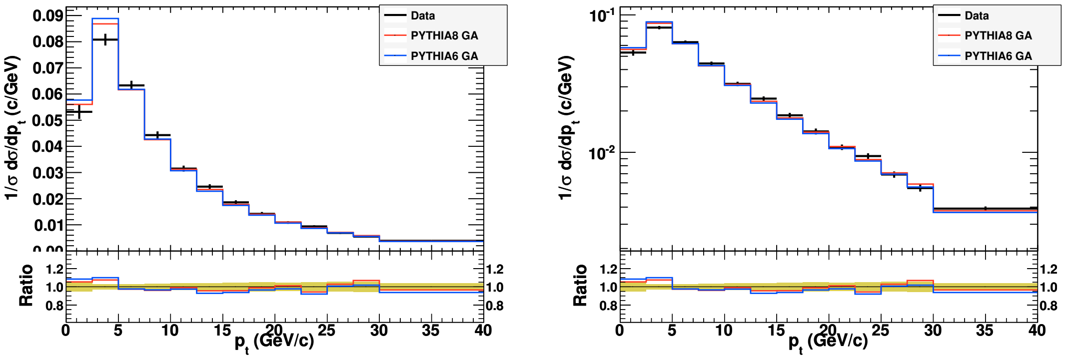


Figure B.9.: DØ $\sqrt{s} = 1960$ GeV $d\sigma/dp_{\mathrm{T}}$ peak region. Right plot is linear and left plot is logarithmic scale [64].

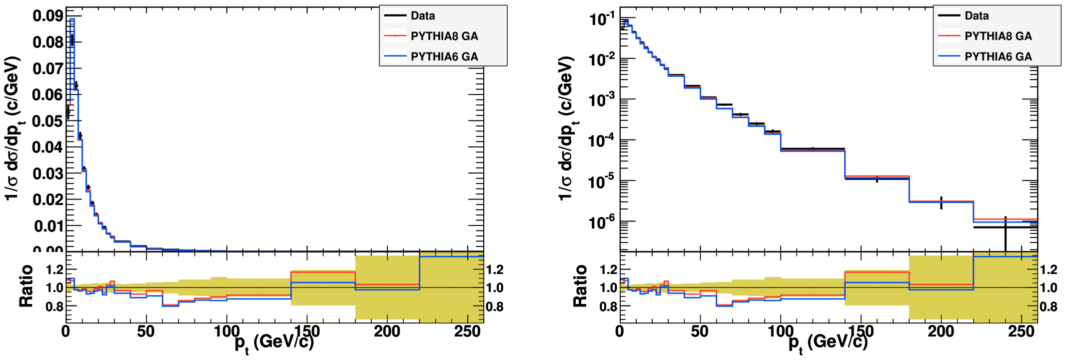


Figure B.10.: DØ $\sqrt{s} = 1960$ GeV $d\sigma/dp_t$. Right plot is linear and left plot is logarithmic scale [64].

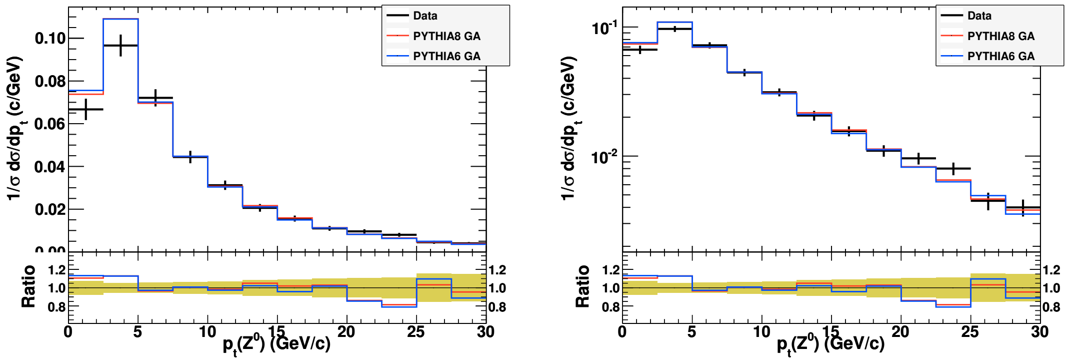


Figure B.11.: DØ $\sqrt{s} = 1960$ GeV Drell-Yan $d\sigma/dp_\mathrm{T}$ for $Z_{p_\mathrm{T}} < 30$ GeV and $|Y_Z| > 2.0$. Right plot is linear and left plot is logarithmic scale [64].

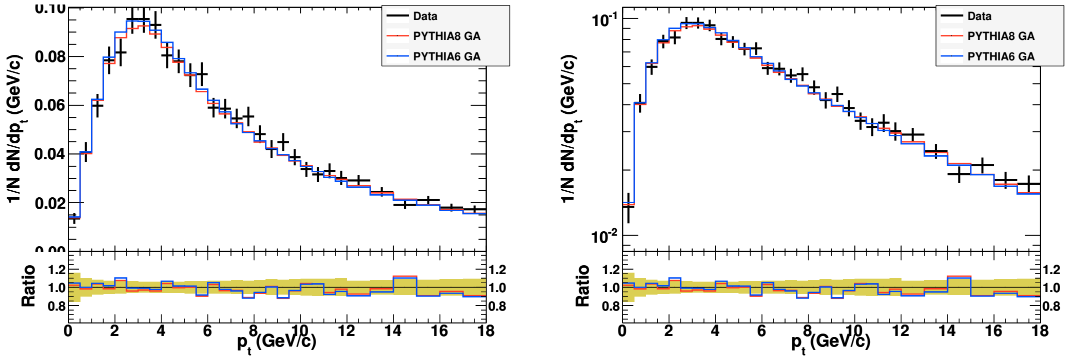


Figure B.12.: CDF $\sqrt{s} = 1800$ GeV normalized Drell-Yan $d\sigma/dp_\mathrm{T}$. Right plot is linear and left plot is logarithmic scale [61].

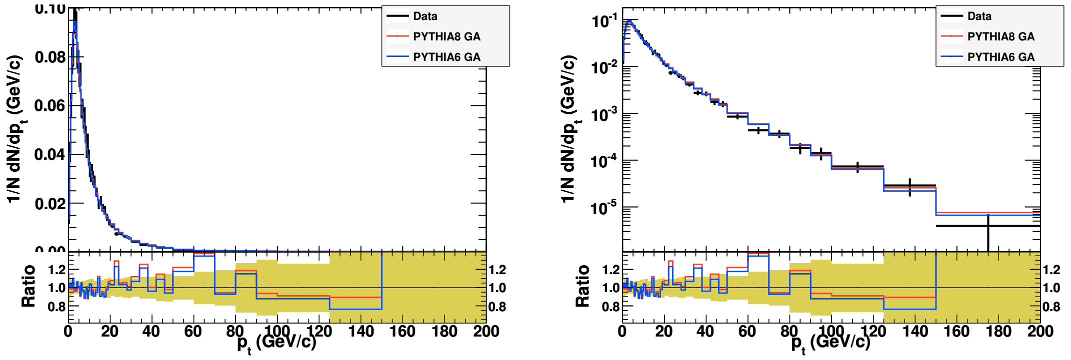


Figure B.13.: CDF $\sqrt{s} = 1800$ GeV normalized Drell-Yan $d\sigma/dp_\mathrm{T}$. Right plot is linear and left plot is logarithmic scale [61].

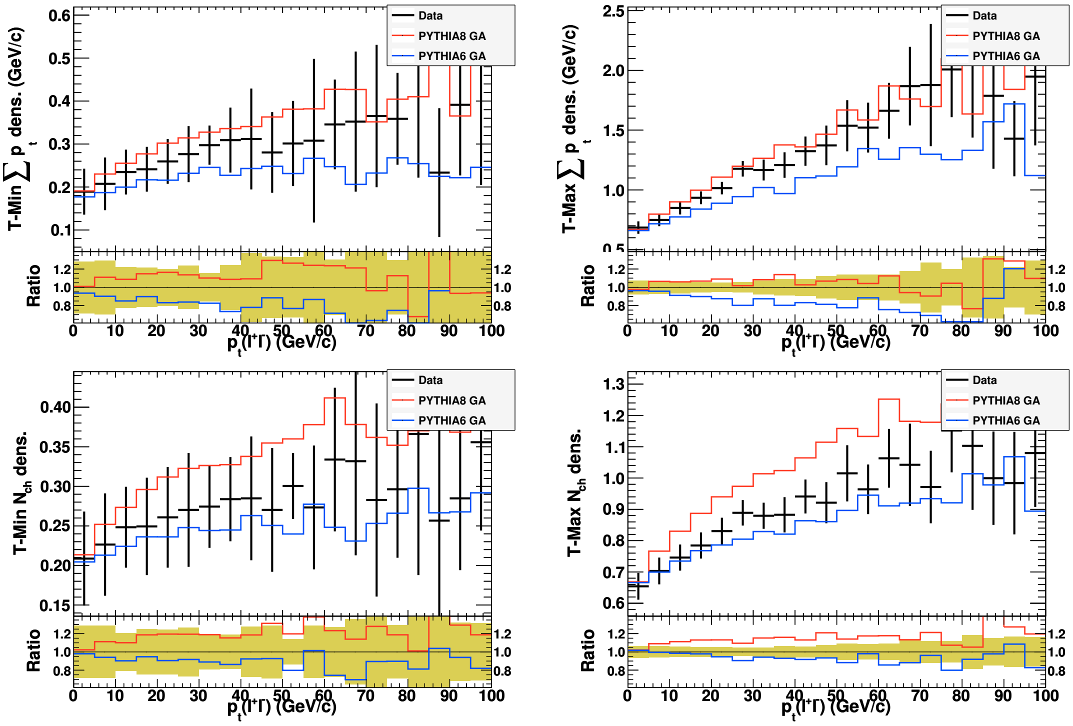


Figure B.14.: CDF $\sqrt{s} = 1960$ GeV Drell-Yan Underlying Event transverse $\sum p_{\mathrm{T}}$ and N_{ch} density. Top row and bottom row show $\sum p_{\mathrm{T}}$ and N_{ch} distributions, respectively, for Trans-Min (left) and Trans-Max (right) regions [62].

C. PYTHIA Tune Comparisons

This appendix contains comparison of tunes found in this thesis with several common PYTHIA 6 and 8 tunes for the data sets that are described in chapter 5. The details of the analyses are given in section 5.1. PYTHIA 6 tunes are selected through PYTUNE subroutine. Table C.1 lists the names of tunes and their identifiers in the PYTUNE subroutine. Peter Skands' tune of PYTHIA 8 is the default tune since version 8.130.

Tune Name	PYTUNE id
DW	103
D6T	109
Perugia-0	320
Professor	329

Table C.1.: PYTHIA 6 tunes and their codes in PYTUNE subroutine.

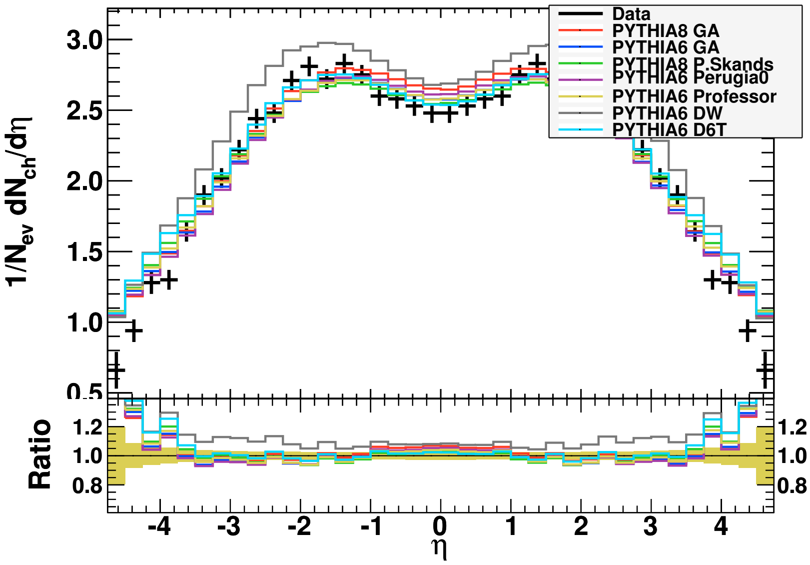


Figure C.1.: UA5 $dN_{\mathrm{ch}}/d\eta$ distribution at $\sqrt{s} = 200$ GeV.

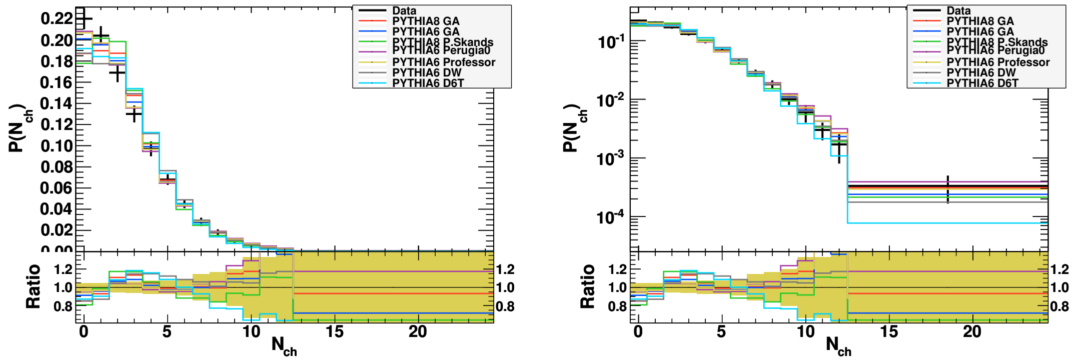


Figure C.2.: UA5 N_{ch} distribution in $|\eta| < 0.5$ at $\sqrt{s} = 200$ GeV. Right plot is linear and left plot is logarithmic scale [66].

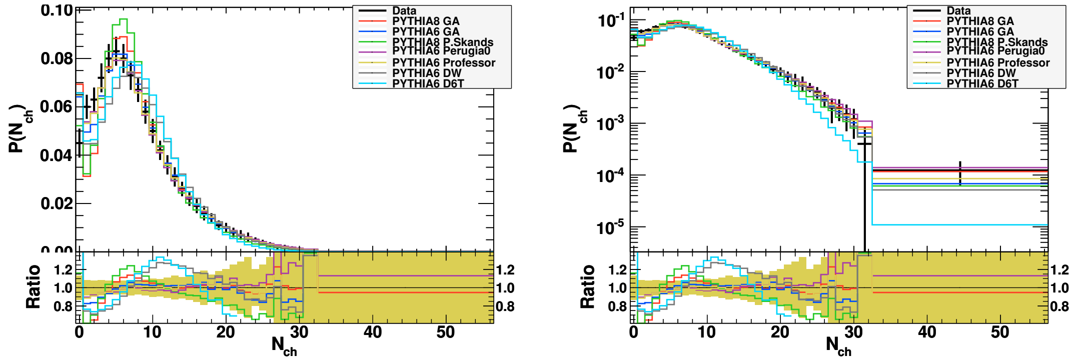


Figure C.3.: UA5 N_{ch} distribution in $|\eta| < 1.5$ at $\sqrt{s} = 200$ GeV. Right plot is linear and left plot is logarithmic scale [66].

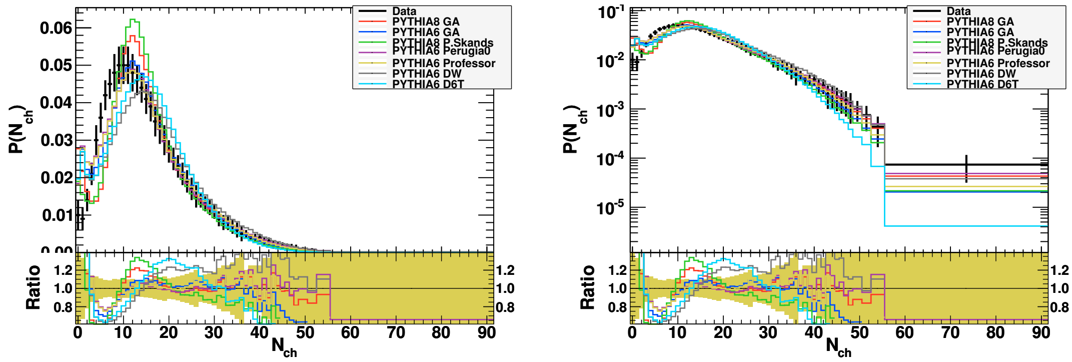


Figure C.4.: UA5 N_{ch} distribution in $|\eta| < 3.0$ at $\sqrt{s} = 200$ GeV. Right plot is linear and left plot is logarithmic scale [66].

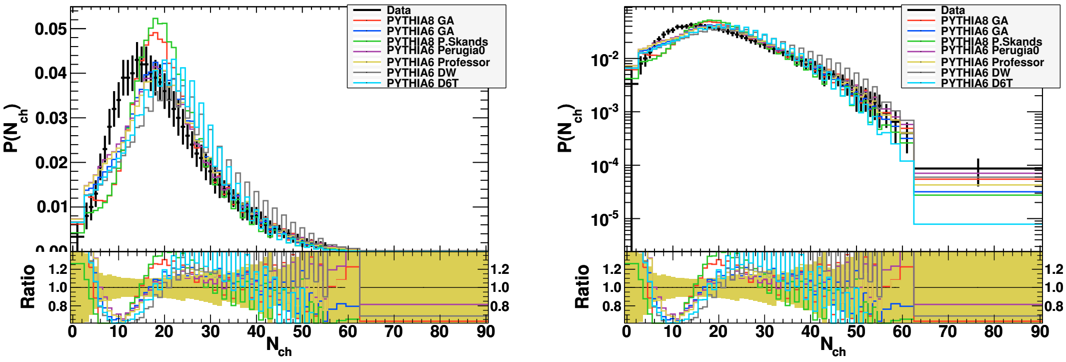


Figure C.5.: UA5 N_{ch} distribution in $|\eta| < 5.0$ at $\sqrt{s} = 200$ GeV. Right plot is linear and left plot is logarithmic scale [66].

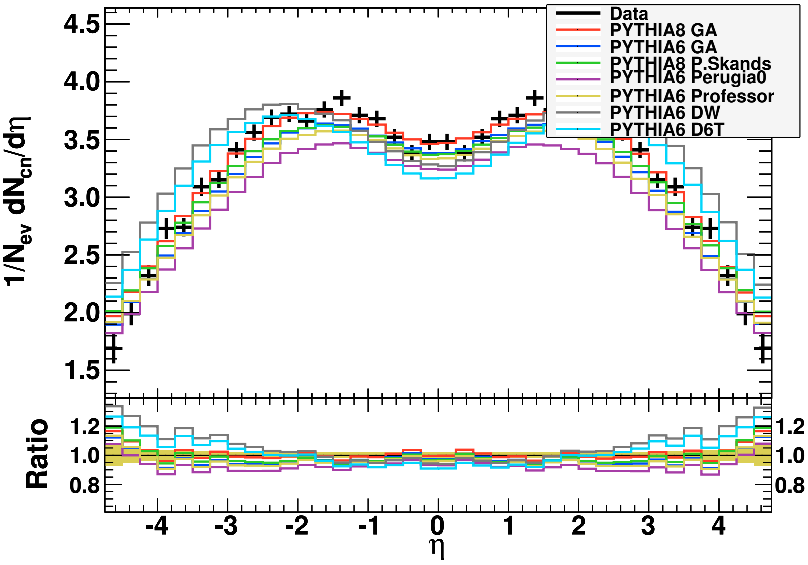


Figure C.6.: UA5 $dN_{\mathrm{ch}}/d\eta$ distribution at $\sqrt{s} = 900$ GeV.

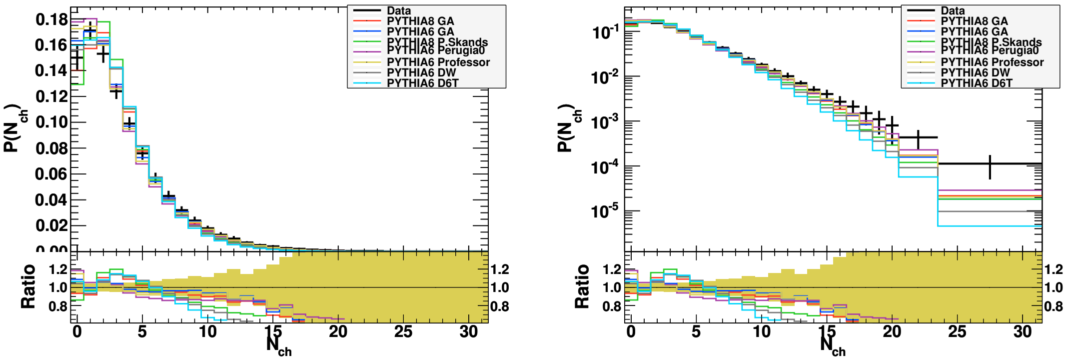


Figure C.7.: UA5 N_{ch} distribution in $|\eta| < 0.5$ at $\sqrt{s} = 900$ GeV. Right plot is linear and left plot is logarithmic scale [66].

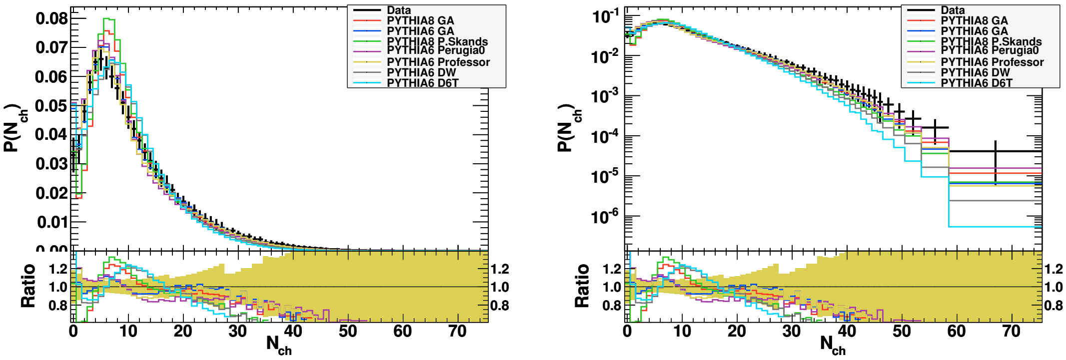


Figure C.8.: UA5 N_{ch} distribution in $|\eta| < 1.5$ at $\sqrt{s} = 900$ GeV. Right plot is linear and left plot is logarithmic scale [66].

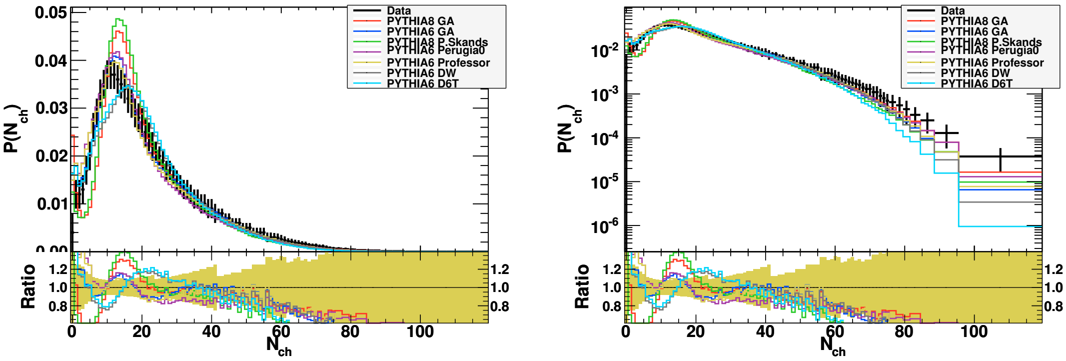


Figure C.9.: UA5 N_{ch} distribution in $|\eta| < 3.0$ at $\sqrt{s} = 900$ GeV. Right plot is linear and left plot is logarithmic scale [66].

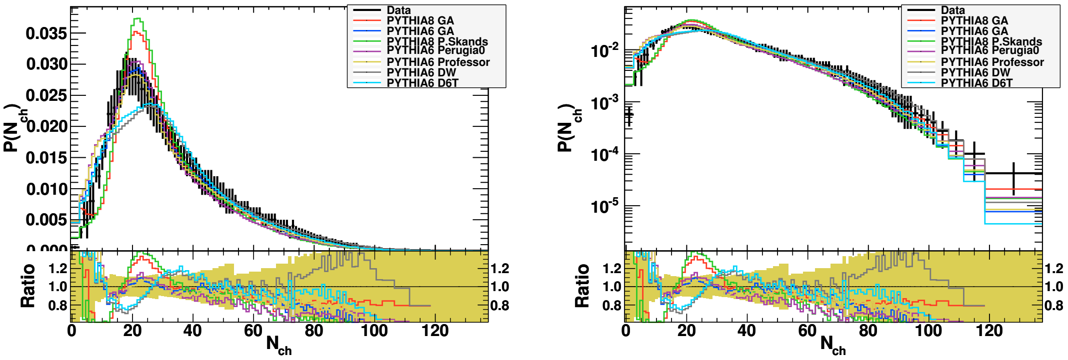


Figure C.10.: UA5 N_{ch} distribution in $|\eta| < 5.0$ at $\sqrt{s} = 900$ GeV. Right plot is linear and left plot is logarithmic scale [66].

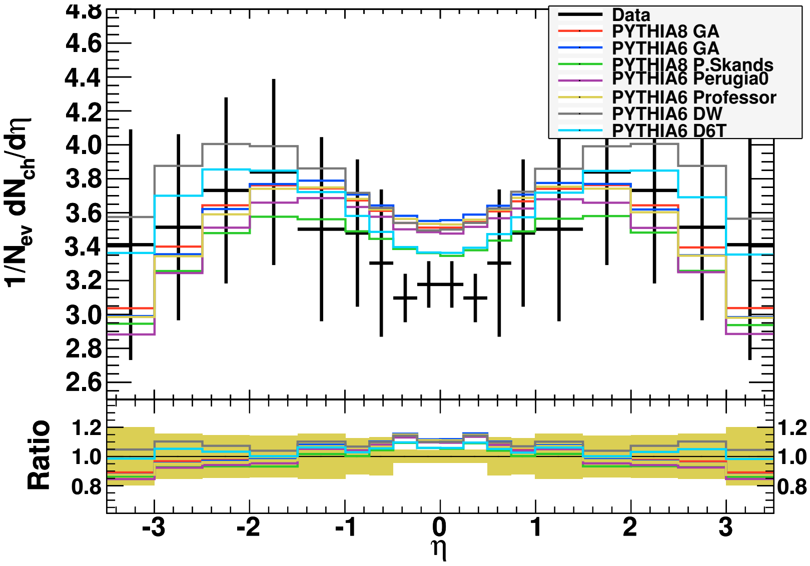


Figure C.11.: CDF $dN_{\mathrm{ch}}/d\eta$ distribution at $\sqrt{s} = 630$ GeV.

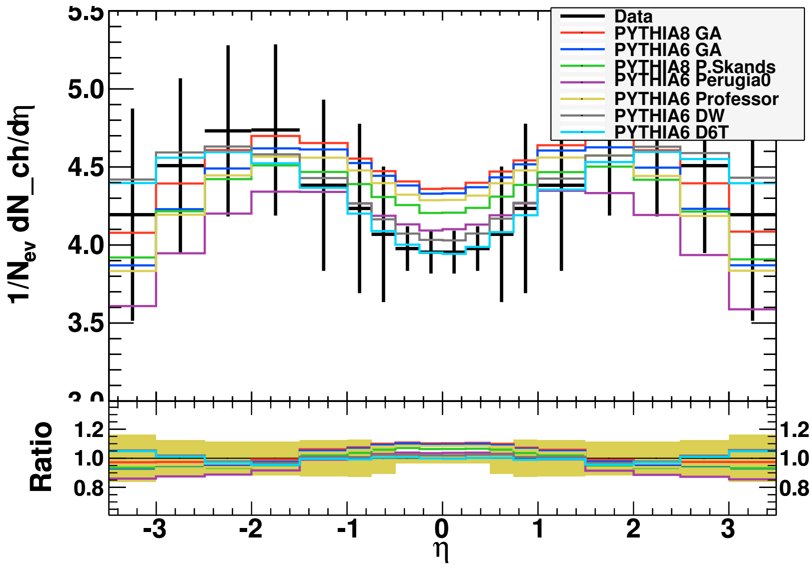


Figure C.12.: CDF $dN_{\rm ch}/d\eta$ distribution at $\sqrt{s} = 1800$ GeV.

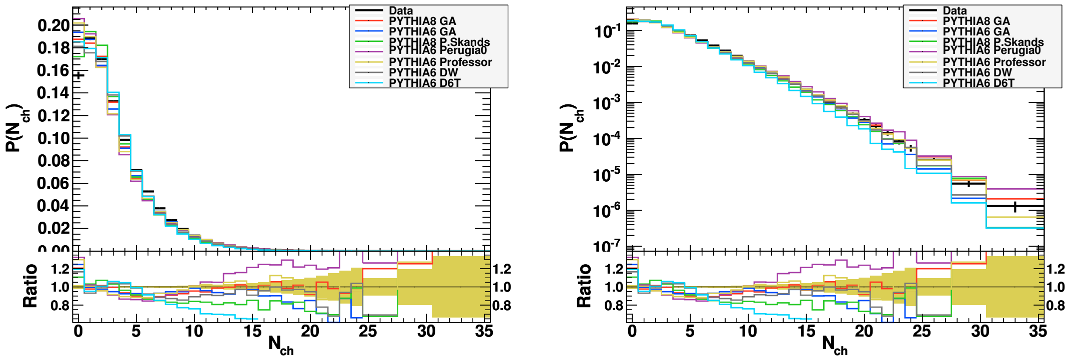


Figure C.13.: CDF $N_{\rm ch}$ distribution in $|\eta| < 1.0$ at $\sqrt{s} = 630$ GeV. Right plot is linear and left plot is logarithmic scale.

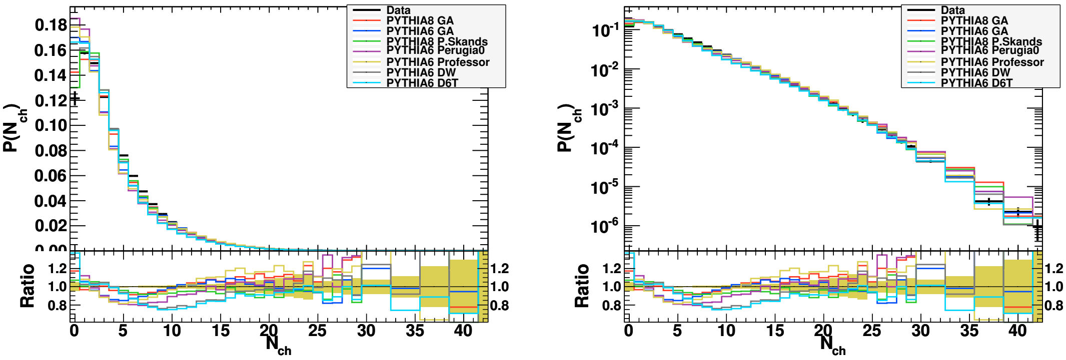


Figure C.14.: CDF N_{ch} distribution in $|\eta| < 1.0$ at $\sqrt{s} = 1800$ GeV. Right plot is linear and left plot is logarithmic scale.

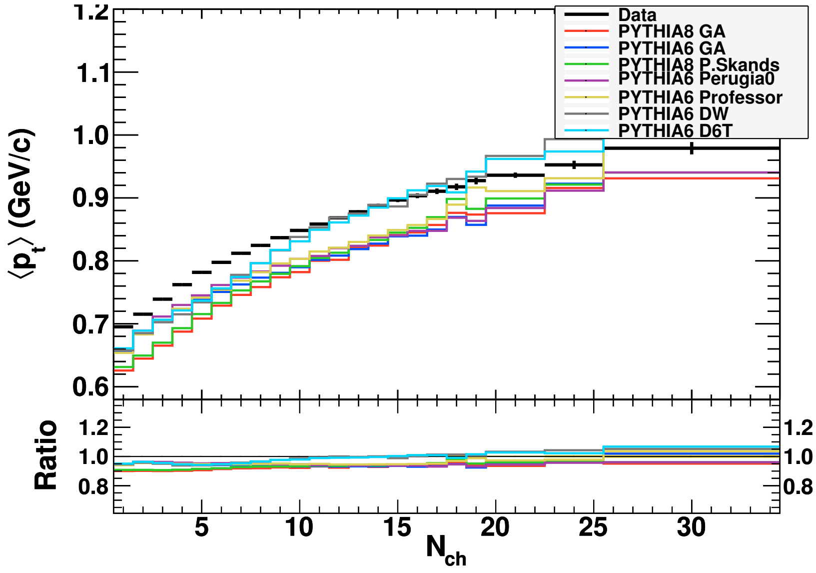


Figure C.15.: CDF $\langle p_{\text{T}} \rangle$ distribution in $|\eta| < 1.0$ at $\sqrt{s} = 630$ GeV.

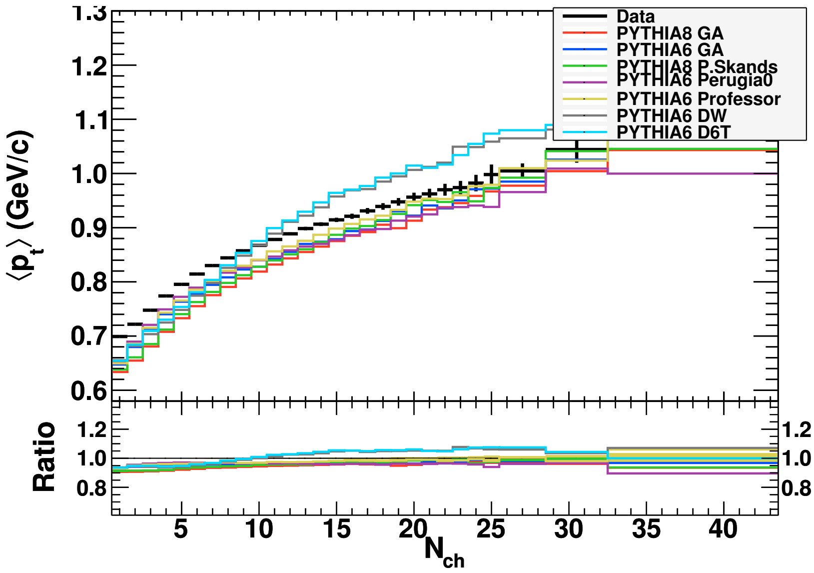


Figure C.16.: CDF $\langle p_{\mathrm{T}} \rangle$ distribution in $|\eta| < 1.0$ at $\sqrt{s} = 1800$ GeV.

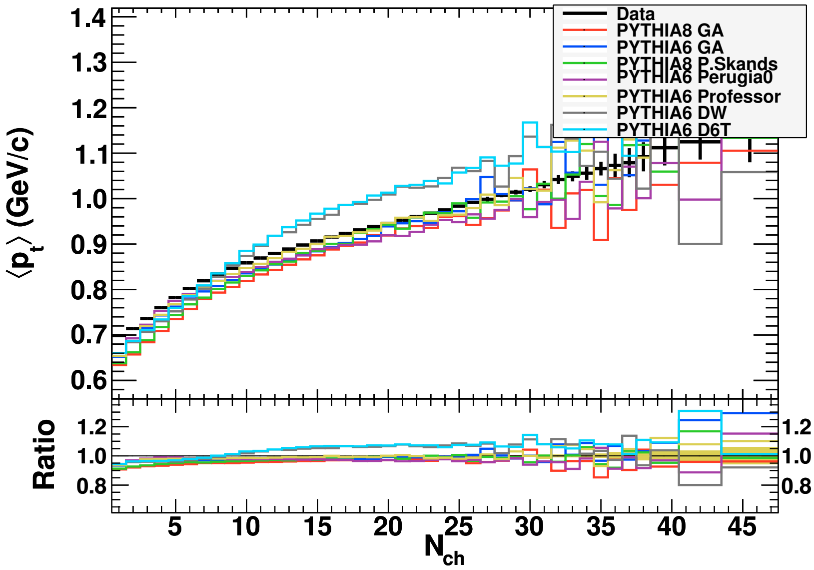


Figure C.17.: CDF $\langle p_{\mathrm{T}} \rangle$ distribution in $|\eta| < 1.0$ at $\sqrt{s} = 1960$ GeV.

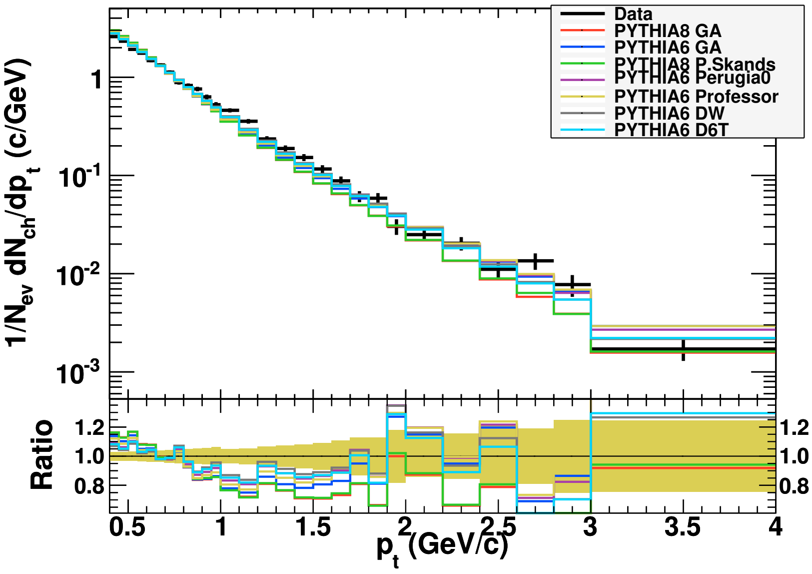


Figure C.18.: CDF $1/NdN_{\mathrm{ch}}/dp_{\mathrm{T}}$ distribution in $|\eta| < 1.0$ at $\sqrt{s} = 630$ GeV.

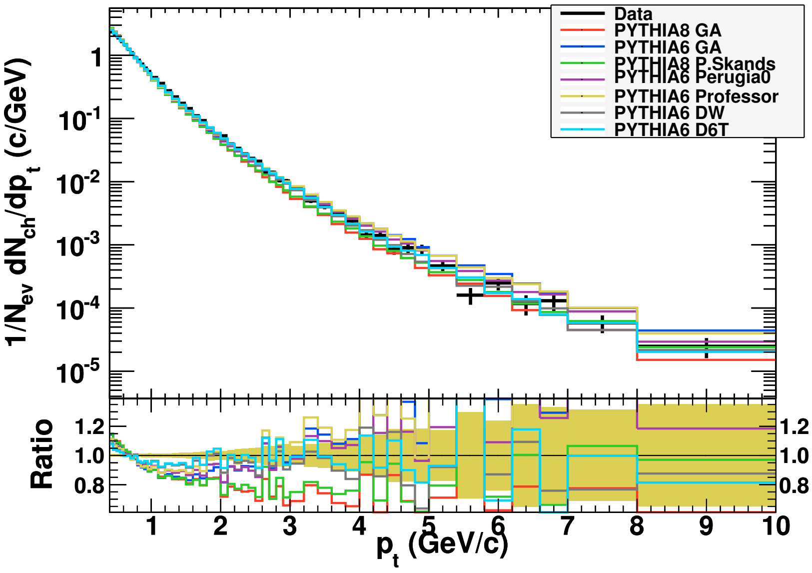


Figure C.19.: CDF $1/NdN_{\mathrm{ch}}/dp_{\mathrm{T}}$ distribution in $|\eta| < 1.0$ at $\sqrt{s} = 1800$ GeV.

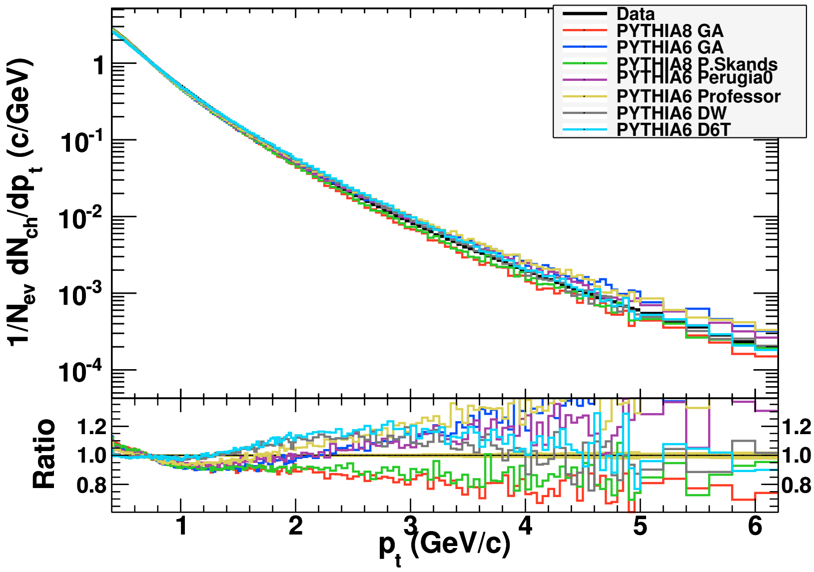


Figure C.20.: CDF $1/NdN_{\text{ch}}/dp_{\text{T}}$ distribution in $|\eta| < 1.0$ at $\sqrt{s} = 1960$ GeV.

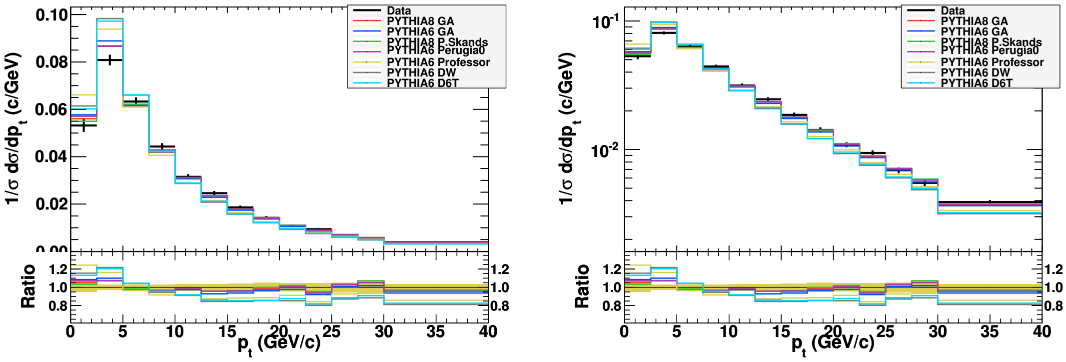


Figure C.21.: DØ $\sqrt{s} = 1960$ GeV $d\sigma/dp_{\text{T}}$ peak region. Right plot is linear and left plot is logarithmic scale [64].

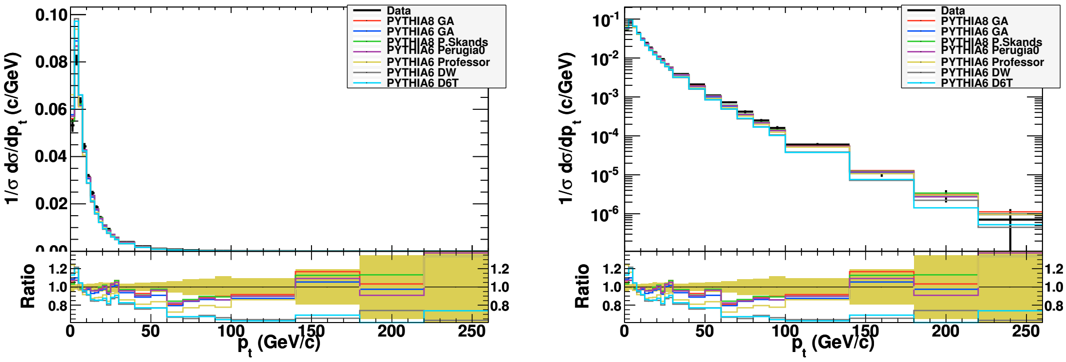


Figure C.22.: DØ $\sqrt{s} = 1960$ GeV $d\sigma/dp_t$. Right plot is linear and left plot is logarithmic scale [64].

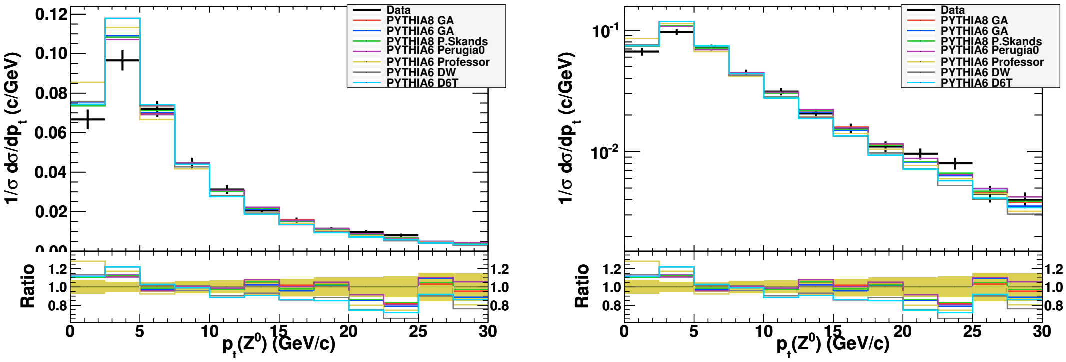


Figure C.23.: DØ $\sqrt{s} = 1960$ GeV Drell-Yan $d\sigma/dp_{\mathrm{T}}$ for $Z_{p_{\mathrm{T}}} < 30$ GeV and $|Y_Z| > 2.0$. Right plot is linear and left plot is logarithmic scale [64].

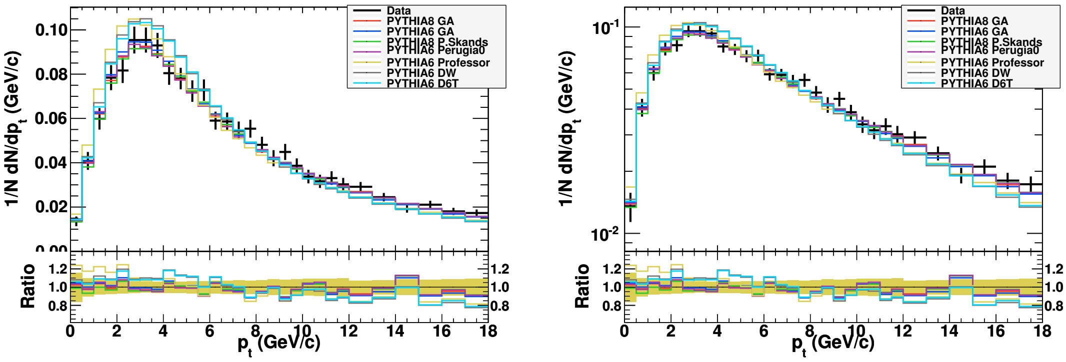


Figure C.24.: CDF $\sqrt{s} = 1800$ GeV normalized Drell-Yan $d\sigma/dp_{\mathrm{T}}$. Right plot is linear and left plot is logarithmic scale [61].

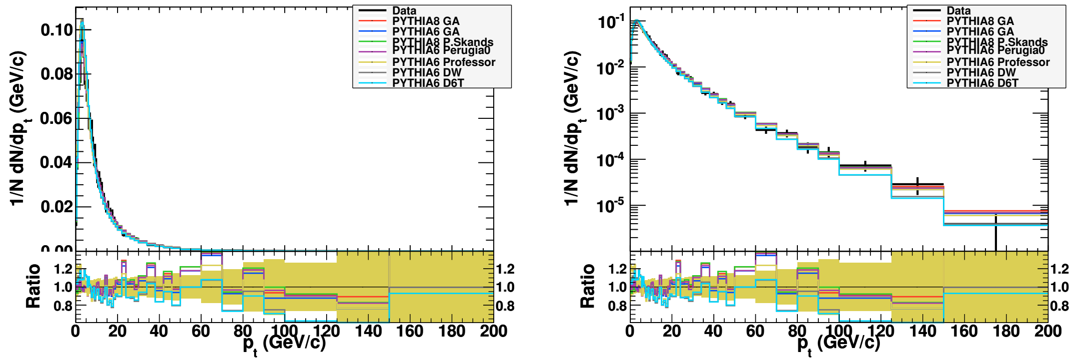


Figure C.25.: CDF $\sqrt{s} = 1800$ GeV normalized Drell-Yan $d\sigma/dp_{\mathrm{T}}$. Right plot is linear and left plot is logarithmic scale [61].

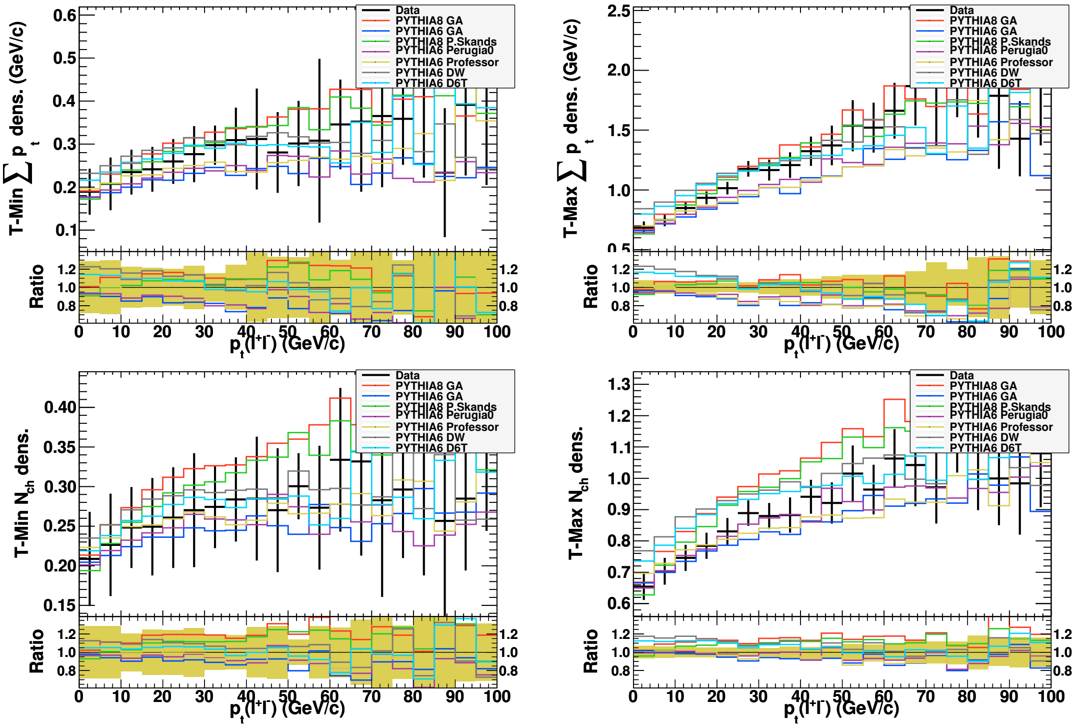


Figure C.26.: CDF $\sqrt{s} = 1960$ GeV Drell-Yan Underlying Event transverse $\sum p_{\mathrm{T}}$ and N_{ch} density. Top row and bottom row show $\sum p_{\mathrm{T}}$ and N_{ch} distributions, respectively, for Trans-Min (left) and Trans-Max (right) regions [62].

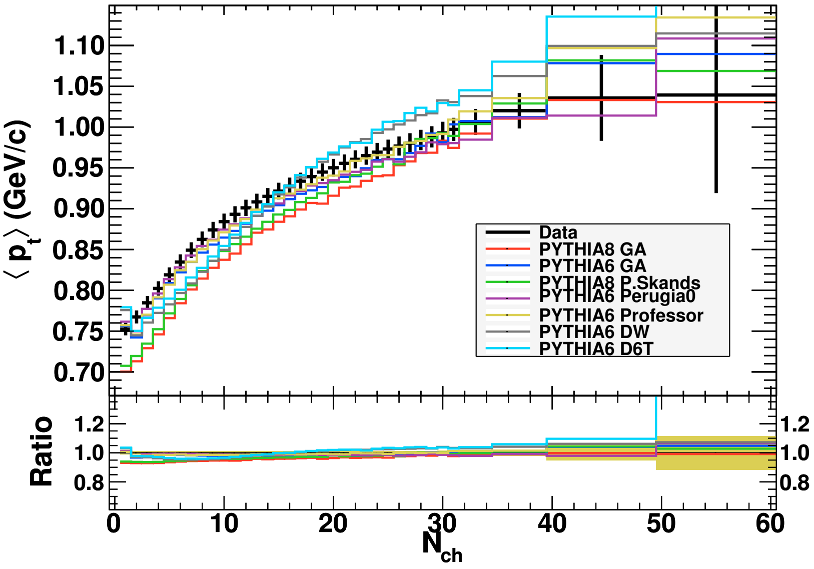


Figure C.27.: ATLAS $\langle p_{\mathrm{T}} \rangle$ distribution for charged particles with $p_{\mathrm{T}} > 0.5$ GeV and $|\eta| < 2.5$ at $\sqrt{s} = 900$ GeV [74].

Bibliography

[1] ATLAS Collaboration. *ATLAS detector and physics performance: Technical Design Report, 2*. Technical Design Report ATLAS. CERN, Geneva, 1999. URL `http://cdsweb.cern.ch/record/391177`.

[2] G. Aad et al. Expected performance of the atlas experiment - detector, trigger and physics. *arxiv*, 0901.0512v4, 2009. URL `http://arxiv.org/abs/0901.0512`.

[3] (ed.) Evans, Lyndon and (ed.) Bryant, Philip. LHC Machine. *JINST*, 2008. doi: 10.1088/1748-0221/3/08/S08001.

[4] G. Aad et al. The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 2008.

[5] R. Adolphi et al. The CMS experiment at the CERN LHC. *JINST*, 2008. doi: 10.1088/1748-0221/3/08/S08004.

[6] A. Augusto Alves et al. The LHCb Detector at the LHC. *JINST*, 2008. doi: 10.1088/1748-0221/3/08/S08005.

[7] G. Anelli et al. The TOTEM experiment at the CERN Large Hadron Collider. *JINST*, 2008. doi: 10.1088/1748-0221/3/08/S08007.

[8] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 2008. doi: 10.1088/1748-0221/3/08/S08002.

[9] G Aad, E Abat, B Abbott, J Abdallah, A A Abdelalim, A Abdesselam, et al. Expected performance of the atlas experiment: detector, trigger and physics. Technical Report arXiv:0901.0512. CERN-OPEN-2008-020, ATLAS Experiment, Geneva, 2009.

[10] M Leyton and J Siegrist. *Minimum Bias Measurements with the ATLAS Detector at the CERN Large Hadron Collider. oai:cds.cern.ch:1200288*. PhD thesis, Berkeley, University of California, Berkeley, Berkeley, 2009. Presented on 13 Aug 2009.

[11] Atlas Collaboration. ATLAS detector and physics performance. Technical design report. Vol. 2. Technical report, CERN-LHC, 1999. CERN-LHCC-99-15.

[12] G. Aad et al. Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics. *arXiv*, 2009. URL `http://arxiv.org/pdf/0901.0512v4`.

[13] Peter Jenni, Marzio Nessi, Markus Nordberg, and Kenway Smith. *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report.* Technical Design Report ATLAS. CERN, Geneva, 2003. URL `http://cdsweb.cern.ch/record/616089`.

[14] I. J. R. Aitchison and A. J. G. Hey. *Gauge theories in particle physics: A practical introduction. Vol. 1: From relativistic quantum mechanics to QED.* IOP, 2003. Bristol, UK: IOP (2003) 406 p.

[15] I. J. R. Aitchison and A. J. G. Hey. *Gauge theories in particle physics: A practical introduction. Vol. 2: Non-Abelian gauge theories: QCD and the electroweak theory.* IOP, 2004. Bristol, UK: IOP (2004) 454 p.

[16] R. Keith Ellis, W. James Stirling, and B. R. Webber. *QCD and collider physics*, volume 8. Cambridge University Press, 1996.

[17] Stephan Narison. QCD as a theory of hadrons (from partons to confinement). *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.*, 17:1, 2002.

[18] C. Amsler et al. Review of particle physics. *Phys. Lett.*, B667:1, 2008. doi: 10.1016/j.physletb.2008.07.018.

[19] John M. Campbell, J. W. Huston, and W. J. Stirling. Hard Interactions of Quarks and Gluons: A Primer for LHC Physics. *Rept. Prog. Phys.*, 70:89, 2007. doi: 10.1088/0034-4885/70/1/R02.

[20] Torbjorn Sjostrand and Maria van Zijl. A Multiple Interaction Model for the Event Structure in Hadron Collisions. *Phys. Rev.*, D36:2019, 1987. doi: 10.1103/PhysRevD.36.2019.

[21] F. Lenz, (ed.), H. Griesshammer, (ed.), and D. Stoll, (ed.). *Lectures on QCD: Applications.* Springer, 1997. Berlin, Germany: Springer (1997) 483 p.

[22] P. D. B. Collins. *An Introduction to Regge Theory and High-Energy Physics.* Cambridge University Press, Cambridge, Mass, 1977. Cambridge 1977, 445p.

[23] Martin M. Block, Kungsik Kang, and Alan R. White. The Current status of high-energy elastic scattering. *Int. J. Mod. Phys.*, A7:4449–4476, 1992. doi: 10.1142/S0217751X9200199X.

[24] K. Hahn and J. Ranft. COMPARING MULTIPARTICLE PRODUCTION WITHIN A TWO COMPONENT DUAL PARTON MODEL WITH COLLIDER DATA. *Phys. Rev.*, D41:1463, 1990. doi: 10.1103/PhysRevD.41.1463.

[25] Klaus Werner. The hadronic interaction model EPOS. *Nucl. Phys. Proc. Suppl.*, 175-176:81–87, 2008. doi: 10.1016/j.nuclphysbps.2007.10.012.

[26] Eugene Levin. An Introduction to pomerons. In *LAFEX International School On High-Energy Physics (LISHEP 98) Session A: Particle Physics For High School Teachers - Session B: Advanced School In HEP - Session C: Workshop On Diffractive Physics*, 1998. URL `http://arxiv.org/abs/hep-ph/9808486v1`.

[27] S. Donnachie, Hans Gunter Dosch, O. Nachtmann, and P. Landshoff. *Pomeron physics and QCD*, volume 19 of *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.*, pages 1–347. Cambridge University Press, 2002.

[28] Otto Nachtmann. Pomeron physics and QCD. In *Ringberg Workshop On New Trends In HERA Physics*, 2003. URL `http://arxiv.org/abs/hep-ph/0312279v1`.

[29] Vincenzo Barone and Enrico Predazzi. *High-Energy Particle Diffraction.* Springer, 2002.

[30] Nicholas Metropolis and Stanislaw Ulam. The monte carlo method. *Journal of the American Statistical Association*, 44(247):335–341, Sep. 1949. doi: 10.2307/2280232. URL `http://www.jstor.org/stable/2280232`.

[31] Torbjorn Sjostrand. Monte Carlo Tools. In *65th Scottish Universities Summer School In Physics: LHC Physics (SUSSP65)*, 2009. URL `http://arxiv.org/abs/0911.5286`.

[32] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[33] E. Boos et al. Generic user process interface for event generators. In *Workshop On Physics At TeV Colliders*, 2001. URL `http://arxiv.org/pdf/hep-ph/0109068v1`.

[34] A. Donnachie and P. V. Landshoff. Total cross-sections. *Phys. Lett.*, B296:227–232, 1992. doi: 10.1016/0370-2693(92)90832-O.

[35] Gerhard A. Schuler and Torbjorn Sjostrand. Hadronic diffractive cross-sections and the rise of the total cross-section. *Phys. Rev.*, D49:2257–2267, 1994. doi: 10.1103/PhysRevD.49.2257.

[36] Sjostrand, Torbjorn. *Pythia8 Web Page*, 2007. URL `http://home.thep.lu.se/~torbjorn/pythiaaux/present.html`.

[37] T. Sjostrand and P. Z. Skands. Multiple interactions and the structure of beam remnants. *JHEP*, 03:053, 2004.

[38] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008. doi: 10.1016/j.cpc.2008.01.036.

[39] A. Capella, U. Sukhatme, C-I Tan, and J. Tran Thanh Van. Dual parton model. *Phys. Rept.*, 236:225–329, 1994. doi: 10.1016/0370-1573(94)90064-7.

[40] Ralph Engel. *PHOJET Manual*, 1.05 edition, 1995. URL `http://www-ik.fzk.de/~engel/phoman5c.ps`.

[41] Ralph Engel. *Hadronic Interactions of Photons at High Energies.* PhD thesis, Universität-Gesamthochschule-Siegen, 1997. URL `http://www-ik.fzk.de/~engel/source/th.ps.gz`.

[42] A. V. Glushkov et al. Muon content of ultra-high-energy air showers: Yakutsk data versus simulations. *JETP Lett.*, 87:190–194, 2008. doi: 10.1134/S0021364008040024.

[43] Klaus Werner, Fu-Ming Liu, and Tanguy Pierog. Parton ladder splitting and the rapidity dependence of transverse momentum spectra in deuteron gold collisions at RHIC. *Phys. Rev.*, C74:044902, 2006. doi: 10.1103/PhysRevC.74.044902. URL `http://arxiv.org/pdf/hep-ph/0506232`.

[44] B. I. Abelev et al. Strange particle production in p + p collisions at s**(1/2) = 200-GeV. *Phys. Rev.*, C75:064901, 2007. doi: 10.1103/PhysRevC.75.064901.

[45] Tanguy Pierog. Forward physics : from SPS to LHC, what can we learn from air showers ? In *44th Rencontres De Moriond On QCD And High Energy Interactions*, 2009. URL `http://arxiv.org/abs/0906.1459v1`.

[46] S. Porteboeuf and K. Werner. Generation of complete events containing very high-p(T) jets. *Eur. Phys. J.*, C62:145–150, 2009. doi: 10.1140/epjc/s10052-008-0840-y.

[47] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput. Phys. Commun.*, 10:343–367, 1975. doi: 10.1016/0010-4655(75)90039-9.

[48] Andy Buckley, Hendrik Hoeth, Heiko Lacker, Holger Schulz, and Jan Eike von Seggern. Systematic event generator tuning for the LHC. *Eur.Phys.J. C65 (2010) 331-357*, 2009.

[49] Andy Buckley, Hendrik Hoeth, Holger Schulz, and Jan Eike von Seggern. Monte Carlo event generator validation and tuning for the LHC. *PoS ACAT08 (2008) 112*, 2009. URL `http://arxiv.org/abs/0902.4403v2`.

[50] Andy Buckley, Hendrik Hoeth, Heiko Lacker, Holger Schulz, and Eike von Seggern. Monte Carlo tuning and generator validation. *1st International Workshop On Multiple Partonic Interactions At The LHC: MPI@LHC 08*, 2009.

[51] P. Abreu et al. Tuning and test of fragmentation models based on identified particles and precision event shape data. *Z. Phys.*, C73:11–60, 1996. doi: 10.1007/s002880050295.

[52] Holger Schulz. *Systematic Event Generator Tuning with Professor*. Diploma thesis, Humboldt-Universität zu Berlin, May 2009. URL `http://projects.hepforge.org/professor/diplomathesis.pdf`.

[53] Durham. Durham reaction database. URL `http://durpdg.dur.ac.uk/HEPDATA/REAC`.

[54] F. Abe et al. The CDF detector: an overview. *Nucl. Instr. Meth.*, A271:387–403, 1988. doi: 10.1016/0168-9002(88)90298-7.

[55] R. Blair et al. The CDF-II detector: Technical design report. Technical report, CDF-II Collaboration, 1996. FERMILAB-PUB-96-390-E.

[56] A. Abulencia et al. Measurements of Inclusive W and Z Cross Sections in p-pbar Collisions at sqrts =1.96 TeV. *J. Phys.*, G34:2457–2544, 2007. doi: 10.1088/0954-3899/34/12/001.

[57] F. Abe et al. Transverse momentum distributions of charged particles produced in $\bar{p}p$ interactions at $\sqrt{s} = 630$ GeV and 1800 GeV. *Phys. Rev. Lett.*, 61:1819, 1988. doi: 10.1103/PhysRevLett.61.1819.

[58] Darin E. Acosta et al. Soft and hard interactions in $p\bar{p}$ collisions at $\sqrt{s} = 1800$-GeV and 630-GeV. *Phys. Rev.*, D65:072005, 2002. doi: 10.1103/PhysRevD.65.072005.

[59] T. Aaltonen et al. Measurement of Particle Production and Inclusive Differential Cross Sections in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev.*, D79:112005, 2009. doi: 10.1103/PhysRevD.79.112005.

[60] F. Abe et al. Pseudorapidity distributions of charged particles produced in $\bar{p}p$ interactions at $\sqrt{s} = 630$ GeV and 1800 GeV. *Phys. Rev.*, D41:2330, 1990. doi: 10.1103/PhysRevD.41.2330.

[61] Anthony Allen Affolder et al. The transverse momentum and total cross section of e^+e^- pairs in the Z boson region from $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev. Lett.*, 84:845–850, 2000. doi: 10.1103/PhysRevLett.84.845.

[62] Rick Field Deepak Kar. Using drell-yan to probe the underlying event in run2 at cdf. *CDFInternal Note 9351*, 2008.

[63] V. M. Abazov et al. The Upgraded D0 Detector. *Nucl. Instrum. Meth.*, A565: 463–537, 2006. doi: 10.1016/j.nima.2006.05.248. URL `http://arxiv.org/abs/physics/0507191`.

[64] V. M. Abazov et al. Measurement of the shape of the boson transverse momentum distribution in $p\bar{p} \to Z/\gamma^* \to e^+e^- + X$ events produced at $\sqrt{s} = 1.96$-TeV. *Phys. Rev. Lett.*, 100:102002, 2008. doi: 10.1103/PhysRevLett.100.102002.

[65] G. J. Alner et al. UA5: A general study of proton-antiproton physics at $\sqrt{s} = 546$-GeV. *Phys. Rept.*, 154:247–383, 1987. doi: 10.1016/0370-1573(87)90130-X.

[66] R. E. Ansorge et al. Charged Particle Multiplicity Distributions at 200-GeV and 900-GeV Center-Of-Mass Energy. *Z. Phys.*, C43:357, 1989. doi: 10.1007/BF01506531.

[67] G. J. Alner et al. Scaling of Pseudorapidity Distributions at c.m. Energies Up to 0.9-TeV. *Z. Phys.*, C33:1–6, 1986. doi: 10.1007/BF01410446.

[68] Mitchell Melanie. *An Introduction to Genetic Algorithms.* MIT Press, 1998.

[69] Matthew Wall. Galib a c++ library of genetic algorithm components. URL `http://lancet.mit.edu/ga/`.

[70] Thomas Weise. *Global Optimization Algorithms - Theory and Application.* Self-Published, second edition, June 26, 2009. URL `http://www.it-weise.de/`. Online available at http://www.it-weise.de/.

[71] Riccardo Poli, William B. Langdon, and Nicholas F. McPhee. *A Field Guide to Genetic Programming.* Lulu Enterprises, UK Ltd, 2008. URL `www.gp-field-guide.org.uk`. ISBN 978-1-4092-0073-4.

[72] J. Bromley et al. HZTOOL: A package for Monte Carlo-data comparison at HERA (version 1.0). In *Future Physics At HERA*, 1996. URL `http://www.desy.de/~heraws96/proceedings/jets/Bromley.ps.gz`. Prepared for Workshop on Future Physics at HERA (Preceded by meetings 25-26 Sep 1995 and 7-9 Feb 1996 at DESY), Hamburg, Germany, 30-31 May 1996.

[73] Edgar Gabriel, Graham E. Fagg, George Bosilca, Thara Angskun, Jack J. Dongarra, Jeffrey M. Squyres, Vishal Sahay, Prabhanjan Kambadur, Brian Barrett, Andrew Lumsdaine, Ralph H. Castain, David J. Daniel, Richard L. Graham, and Timothy S. Woodall. Open MPI: Goals, concept, and design of a next generation MPI implementation. In *Proceedings, 11th European PVM/MPI Users' Group Meeting*, pages 97–104, Budapest, Hungary, September 2004.

[74] Atlas Collaboration. Charged-particle multiplicities in pp interactions at $\sqrt{s} = 900$ GeV measured with the ATLAS detector at the LHC. *Phys. Lett. B.*, 2010. URL `http://arxiv.org/abs/1003.3124v1`.

[75] The CMS Collaboration. Transverse momentum and pseudorapidity distributions of charged hadrons in pp collisions at sqrt(s) = 0.9 and 2.36 TeV. *JHEP*, 02:041, 2010. doi: 10.1007/JHEP02(2010)041.

[76] B. I. Abelev et al. Systematic Measurements of Identified Particle Spectra in pp, d^+ Au and Au+Au Collisions from STAR. *Phys. Rev.*, C79:034909, 2009. doi: 10.1103/PhysRevC.79.034909.

[77] ALICE Collaboration. First proton–proton collisions at the LHC as observed with the ALICE detector: measurement of the charged particle pseudorapidity density at sqrt(s) = 900 GeV. *Eur. Phys. J.*, C65:111–125, 2010. doi: 10.1140/epjc/s10052-009-1227-4.

[78] C. Albajar et al. A Study of the General Characteristics of $p\bar{p}$ Collisions at $\sqrt{s} = 0.2$-TeV to 0.9-TeV. *Nucl. Phys.*, B335:261, 1990. doi: 10.1016/0550-3213(90)90493-W.

[79] A. M. Rossi et al. Experimental Study of the Energy Dependence in Proton Proton Inclusive Reactions. *Nucl. Phys.*, B84:269, 1975. doi: 10.1016/0550-3213(75)90307-7.

[80] ATLAS Collaboration. Athena framework. URL `https://twiki.cern.ch/twiki/bin/view/Atlas/AthenaFramework`.

[81] S. Agostinelli et al. GEANT4: A simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003. doi: 10.1016/S0168-9002(03)01368-8.

[82] G. Piacquadio, K. Prokofiev, and A. Wildauer. Primary vertex reconstruction in the ATLAS experiment at LHC. *J. Phys. Conf. Ser.*, 119:032033, 2008. doi: 10.1088/1742-6596/119/3/032033.

[83] Doris Burckhart-Chromek, Voica A. Radescu, and Sami Kama. Requirements for operational monitoring display. Atlas internal note., 2007. URL `https://edms.cern.ch/file/847221/1/OMDRequirementsRev37.pdf`.

List of Figures

List of Tables

Acknowledgment

There are several people that I want to thank. First and foremost I would like to thank my supervisor, Dr. Klaus Mönig, for his kind support, encouragement, guidance, knowledge and patience; all of which made this dissertation possible. Prof. Hermann Kolanoski explained subjects that I didn't understand, asked valuable questions which lead me to understand the subject better and went through carefully over my thesis, making important comments and corrections. Prof. Torbjörn Sjöstrand responded to my emails and answered my questions about the PYTHIA kindly and quickly and spared some of his time at CERN to explain the details of PYTHIA. Dr. Müge Karagöz read the thesis, corrected many mistakes that I did and suggested various clarifications. Dr. Tanguy Pierog supplied the EPOS code and provided assistance about it whenever I needed. Andreas Haupt, Götz Waschk, Stephan Wiesand and many others from DESY Zeuthen computing center made necessary changes in the batch systems for me to be able to use my program at full potential. While I worked at CERN, Serguei Kolos helped me understand the online monitoring framework and gave me crucial advices in design of the OMD and computing. I would like to thank them all for their help on this thesis.

I want to thank my colleagues, Judita Mamuzic, Michael Leyton and Conrad Friedrich for nice conversations, from daily chit-chat to serious physics discussions and Regina Kwee for her help with abstract in German and friendly discussions including those on minimum bias.

Finally I want to thank my wife for she took care of everything in my life so that I could concentrate on my dissertation. Without her, preparing this dissertation would be much harder for me.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 23.03.2010 Sami Kama